Published on *Marine Metadata Interoperability* (https://marinemetadata.org)

MMI Guides

Welcome to the MMI Guides

The Marine Metadata Interoperability (MMI) Guides are intended to

- help researchers and data managers learn about metadata;
- encourage best practices in metadata development, interoperation, and distribution;
- foster community involvement in this process; and
- facilitate technical understanding of the role and importance of marine metadata.

Experienced scientists, technologists, and publishers have distilled the more complex aspects of marine metadata into these documents.

To start, read How to Use the Guides [1] on getting the most out of the material and how to cite the guides.

If you need additional information, or if you'd like to contribute your experiences with marine metadata, contact us by e-mail [2].

- MMI Guides Editorial Group [3]

Citing the Guides

To cite an individual guide, please follow the citation example found on that guide page. To cite the MMI Guides as a whole, please use: Stocks, K.I., Neiswender, C., Isenor, A.W., Graybeal, J., Galbraith, N., Montgomery, E.T., Alexander, P., Watson, S., Bermudez, L., Gale, A., Hogrefe, K., 2010, accessed [date, e.g. June 13, 2010]. *The MMI Guides: Navigating the World of Marine Metadata*. http://marinemetadata.org/guides [4]

How to Use the Guides

The Marine Metadata [5] Interoperability [7] (MMI [9]) Guides are organized into two parts: major topic areas (Introduction to Metadata [10], Metadata Standards [11], and Vocabularies [12]) and supporting pages (Case Studies, Additional Resources, and Glossaries [13] of Terms and Acronyms). Each web page is considered to be one guide.

Within each topic area, the top-level guides are the most introductory and assume little previous knowledge. The lower-level guides in each section are more detailed and assume more advanced knowledge.

The Introduction to Metadata guide contains overviews of many topics and is a good starting place for researchers new to the world of metadata. More advanced users may find the Table of Contents [15] easier to quickly access information on particular topics.

Terminology

Within the guides, terms in the glossary [13] are highlighted on each page the first time they appear. Move your cursor over a highlighted word to see the definition in a pop-up box. Click on the word to go to the glossary page, which contains the definition as well as a link to any related pages in the guides for more detailed information about the topic.

Other Tools and Navigation

- The right side of each page contains a list of glossary terms that are used in the article along with relevant links.
- The bottom of the page contains the suggested citation format for each guide and a feedback link for submitting questions and suggestions.
- The left side of each page contains the main navigation tools for the MMI site as well as the table of contents for the MMI Guides.

Citing the MMI Guides

- To cite an individual guide, follow the citation example found on that guide page.
- To cite the MMI Guides as a whole: Stocks, K.I., Neiswender, C., Isenor, A.W., Graybeal, J., Galbraith, N., Montgomery, E.T., Alexander, P., Watson, S., Bermudez, L., Gale, A., Hogrefe, K., 2010. *The MMI Guides: Navigating the World of Marine Metadata*, accessed [date, e.g., July 21, 2010], http://marinemetadata.org/guides [4].

How to cite this Guide

Neiswender, C. 2010. "How to Use the Guides." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org /guides/newuserrecs. Accessed June 23, 2014.

Table of Contents

Introduction to Metadata

Introduction to metadata and the MMI Guides on metadata

The most common definition of metadata [5] is that metadata are data about data. Metadata describe the who, what, when, where and how of a resource. Implicit in this definition is the purpose of metadata: to attach information to data, so they can be discovered and used. See the guide, Definition of Metadata [6] for a more detailed description of this term.

In today's research environment, creation of metadata is becoming a requirement for practical use of research observations and results. You need metadata if you want to

- Find data from other researchers to support your research.
- Use the data that you do find.
- Help other professionals to find and use data from your research.
- Use your own data in the future when you may have forgotten details of the research.

So whether you are retrieving or distributing data, understanding the principles and practice of metadata works to your benefit.

In this section you will learn basic metadata concepts and approaches. Definition of Metadata [6], plus the lower-level guides on Metadata Classifications [19] and Types of Vocabularies, [103] are where commonly used terms are defined.

The next three sections—The Importance of Metadata [20], Metadata Interoperability [8], and Machine Readability— [21]describe rationales for metadata.

Getting Started with Metadata [22] is where you start laying out the steps to develop and implement metadata for your project. Writing Good Metadata [24] provides a quick checklist of characteristics that your metadata should have. And finally, Some Examples of Metadata [25] connects to real-world examples of completed metadata.

In the later sections of the guides, metadata standards [11] and specifications [104] are explored, as well as controlled vocabularies [12]. Controlled vocabularies, which include thesauri [105], gazetteers [106], and ontologies [107], are critical tools for creating interoperability between—and standardization within—datasets.

How to cite this Guide

Neiswender, C. 2010. "Introduction to Metadata." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdataintro. Accessed June 23, 2014.

Definition of Metadata

Description of metadata, including definition, the difference between metadata and data, and metadata examples

Metadata [5] are used to describe data or information. In environmental sciences like oceanography, metadata describe the information that scientists collect and informs users about the characteristics and history [108]

The record of how a particular value or record came to be. Provenance can include things like when, by whom, and how the item was created or modified.

of a data set or data item—including methodological, temporal and spatial information.

The word metadata is sometimes used in a singular form (metadata is). We use the plural (metadata are). Both are in common usage, though in the sciences it's typically used in the plural.

The National Information Standards Organization (NISO [109]) defines metadata as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource." The World Wide Web Consortium (W3C [110]) defines metadata as "machine understandable information for the web." The Federal Geographic Data Committee (FGDC [111]) defines metadata as describing, "the content, quality, condition, and other characteristics of data." Put simply, metadata are data about data. They provide context for research findings, ideally in a machine-readable [112] format. Once published, metadata can enable discovery [113] of data via electronic interfaces and enable correct use and attribution of your findings.

The results of data collection are generally objects—photos, spreadsheets, maps, graphs, data files, etc. These objects are useful but generally do not contain information about how, where, or by whom the data was collected. The data object alone is difficult to interpret and use. However, if you provide the right descriptive information, your data become much more useful. The additional information might include things like latitude and longitude, date collected, precision of the measurement, person to contact with questions about the data, or type of equipment used. This context, or descriptive data, is the metadata.

The Difference between Metadata and Data

Metadata describe a data set sufficiently to permit searching and using the data. However, it is not always clear if a particular piece of information should be classified as data or metadata. Some information, such as geographic coordinates of observations, can be classified as both data and metadata. The distinction between metadata and data depends on the context and the needs of a given application or user.

Briefly stated, any data that are required to make other data useful or searchable can be called metadata. Again quoting the NISO guide, *Understanding Metadata*, "Metadata is key to ensuring that resources will survive and continue to be accessible into the future." To illustrate this important distinction, consider a conductivity-temperature-depth (CTD [114]) profile [115] where a single temperature measurement may be lost without significantly degrading the value of the profile. The loss of positional information from a metadata record, however, renders the data almost useless.

Metadata Structure

Metadata do not have to be in any particular format to qualify as metadata—notes scribbled onto a post-it note and stuck onto your computer monitor can be metadata, though not particularly useful ones. Later guides discuss the importance of machine readability [21] for metadata, and of metadata standards and specifications [11]. Here, we introduce you to the terms commonly used to describe metadata: element [117] and value.

Elements refer to the categories of metadata used. They can also be referred to as properties, or more informally, as fields. Values [119] refer to the actual information filled into an element. Using the above examples, latitude would be an element, and +32.5 might be a value for that element from a particular data record. Similarly, core type would be an element, and piston core and megacore would be values for that property.

Examples of Metadata

Television programming provides a simple example of metadata. When you turn on a television and want information about the next show, you will probably go to an index of television shows. You may consult the TV listings in a newspaper or in *TV Guide*, or you may view on-screen program information. The listings you look at contain data (show title, type of show, time, or plot summary) about other data (the television broadcasts themselves).

Scientific examples are more complex, but the same concept applies. The notes written by scientists about their experiments—in lab notebooks, log books, or other documents—are metadata (information) about data (the results of the experiment). The notes describe characteristics of the experiment.

Oceanographic examples

Data: Photo of a newly discovered species of fish
Metadata: Location of discovery (latitude, longitude, and depth), other fish in the area, salinity of the water, quantity discovered (school, single fish, two or three individuals), etc.
Data: Meteorological Measurements
Metadata: Location of readings (latitude, longitude, and height), instrumentation used to collect data, units, processing done to measurements

Data: Sediment Core Record
Metadata: Location of discovery (latitude, longitude and depth), description of stratigraphy, length, type of coring device

Notice that different types of data require different types of descriptive information. However, there are some standard fields that should always be included (for example, location and date collected).

How to cite this Guide

Graybeal, J., Neiswender, C., Stocks, K., Bermudez, L., Galbraith, N., Watson, S., Miller, S.P. 2010. "Definition of Metadata." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdataintro/mdatadefined. Accessed June 23, 2014.

Metadata Classifications

Description of the main ways metadata are categorized, classified, or organized into types.

In developing data systems, a number of categories have been used for metadata [5]. This guide explains a few of these and some of their strengths and weaknesses.

Many of them are not particularly well defined and represent messy and partially overlapping ways to organize metadata. They are presented here for two reasons. First, so that you will know their meaning if you run across these terms in other metadata-related reading. Second, because in classifying metadata in different ways they highlight the different uses of metadata, which can be valuable to consider when developing your own metadata approach.

Metadata Classification Techniques

- **Syntactic vs Semantic**
  The structure of the data (syntax) as opposed to their meaning (semantic).
- **Use vs Search**
  Metadata required for someone to make appropriate use of the data, as opposed to metadata required for finding the data.
- **Static vs Dynamic**
  Whether the metadata change through time.
- **By Functional Category**
  Includes six different functions performed by metadata.

Syntactic vs Semantic

Syntactic metadata describe what the data look like and how they are organized. Semantic metadata describe what they mean. Semantic data are often considered to be human-oriented rather than machine-usable, but that seems to be an assumption, and not required by the term itself.

Syntactic fields often include the unique variable name, data type (integer, float, etc., including sizes), file format, and units of measurement. Note that the first and last of these certainly have semantic meaning, even if their primary use is for labeling or identification.

Semantic fields are often more descriptive, such as long name, definition, comments, and copyright. For example, the information that a field labeled "SST [121]" holds sea surface temperature measurements is semantic metadata. Most semantic fields would be more widely useful if they followed agreed-upon conventions and terminology. The increasing use of ontologies [107] will likely push semantic content much more into a machine-readable [112] realm.

Use vs Search

Search metadata, also known as discovery metadata [113], include information that would help a person decide if there were things of interest in a data set or which search keywords to use if they were using a data portal. An observation type such as multibeam bathymetry is an example of helpful search metadata, especially when managed by a system of controlled vocabularies [122]. Search metadata might also be latitude and longitude bounds, so that a computer or a person could know if the data fell within an area of interest.

Usage (use) metadata helps a computer or a person to understand or process the data. Typical use metadata would be calibration parameters, units, and precision information. Use metadata often overlap with syntactic metadata, though they are not synonymous. Usage metadata [123] labels need to be unique to be of value [119] for processing the data, while syntactic data may not.

Based on typical definitions of the terms, the distinction between use and search metadata can be unclear. Some or all search metadata may be automatable, that is, represented in ways that are meaningful to the applications processing and used by that software. Indeed, this will be necessary to facilitate widespread data mining. Furthermore, some use metadata will be of interest to people searching for data, even though it is more oriented toward computer applications.

When designing a metadata approach, it is important to consider both the terms and characteristics that people or systems will need to search for and to find your data, and also the details that people and systems would need to know to use them.

Static vs Dynamic

Static metadata are not expected to change much over the life of the data they describe, even as the data evolve. Conversely, dynamic metadata are a function of the contents of the data, so as a data set evolves, dynamic metadata change. In reality, even static metadata may have to be changed if, for example, incorrect information was captured and the error discovered later.

An interesting special case involves the seeding of metadata prior to the arrival of data themselves. Metadata captured before data arrival are implicitly static and can be associated with that data permanently, possibly as part of an automated process embedded in the data stream. Metadata captured after data arrival imply some other process for entering that information.

When planning your metadata process, it is worth considering which metadata you expect to be persistent through time, and which will need to change. And you will need to determine the processes for updating dynamic metadata, and for handling an unexpected need to adjust or correct static metadata.

By Functional Category

In their 2006 paper (see reference below), Ganesan Shankaranarayanan and Adir Even propose six types of metadata based on their function:

- Infrastructure metadata describes the components of the computer systems, such as hardware, operating systems, networking and database servers. It is primarily used for system maintenance.
- Model metadata (the data dictionary [124]) describes the modeling of data into entities and their relationships [125] (e.g., tables and column headers). It includes conceptual, logical, and physical descriptions as well as semantic information, such as terms used and how they relate to other terminology in the system.
- Process metadata provides information on how data are generated and the changes they undergo from source to target.
- Quality metadata includes both a description of the physical size (number of records, bytes) as well as quality measurements, such as the accuracy and completeness of the data.
- Interface (delivery and reporting) metadata captures how the data is used, such as where and how much data are delivered (e.g., downloaded from an online system) and in what formats. It can also include how the data are used in derivative products like reports.
- Administration metadata includes information on users, security, and access privileges to data and applications.

Classification Summary

The classifications above represent frameworks to consider in designing a metadata system. When designing your system, it is important to understand what kind of data you are dealing with, what kinds of questions you need to answer with your metadata, and which distinctions above are likely to be relevant to your system and which are not.

Finally, the categories above illustrate the importance of precise terminology when collaborating on a design. Be sure that the data system's developers are using *search metadata* to mean the same thing you are. Lists of metadata fields [117], and the user queries they will enable, are helpful tools to ensure agreement and understanding.

References

Shankaranarayanan, G. and A. Even. The Metadata Enigma, *Communications of the ACM*, Vol. 49, No. 2, pp 88–94, February 2006.

How to cite this Guide

Graybeal, J., Stocks, K., Miller, S.P. 2010. "Metadata Classifications." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdataintro/mdatadefined/howclassified. Accessed June 23, 2014.

The Importance of Metadata

Discussion of the value of metadata, and how it benefits research

If you have ever wanted to use someone else's data, such as for a meta-analysis or for comparison with your own data, you may have had a difficult time finding, deciphering, and using that data. You may have found that critical information about the data is missing. Worse, there may have been no clear contact information where you could send questions.

Regardless of the source of the data, whether it is a colleague next door or an international data repository, there is data-related information that you, as a scientist, expect and need. You need to know its storage location, format, and time and location of the measurements.

You also need to know the originator, contact information, initial collection intent, instruments used, processing methods and algorithms, and quality of the data. In fact, this information is required even when using your own data, though you may keep much of it in your head instead of in structured metadata [5].

Metadata are intended to provide all of this data-related information. Metadata help you find, access, understand, and use the data in the following ways. Metadata:

- **Make the data easier to manage.**
  Data managers don't have to repeatedly answer the same questions about processing methods, data quality, etc. Effectively, the metadata reduce duplication of effort, facilitate sharing of reliable information, streamline workload, and publicize or document work.
- **Make data more useful to more people.**
  Those attempting to use the data have their data-related questions answered by the metadata. For the data user, the metadata make it possible to effectively find and use valuable data from other research projects.
- **Promote human and machine readability.**
  Metadata provide data-related information in a form that can be read by a human or used in automated processing. Metadata become far more valuable when maintained in standard ways and stored in a computer so they may be searched by and distributed to others.
- **Fosters collaboration.**
  Metadata will help raise awareness of the quality of data and activities in your organization. In turn, this could help establish new collaborations that further utilize your valuable data assets.
- **Avoids costly duplication.**
  When data are easily found and accessed, the risk of accidental duplication of effort is reduced—scientists are less likely to reproduce research that has already been carried out.

Benefits of having quality metadata are further outlined in the NISO [109] guide, *Understanding Metadata*. As stated in the NISO guide [126], *"Metadata is key to ensuring that resources will survive and continue to be accessible into the future."*

How to cite this Guide

Hogrefe, K., Stocks, K. 2010. "The Importance of Metadata." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdataintro/mdataimportance. Accessed June 23, 2014.

Metadata Interoperability—What Is It, and Why Is It Important?

Introduction to interoperability - the exchange of metadata between computer systems

Metadata interoperability is the ability of two or more information systems to exchange metadata with minimal loss of information. This does not address data compatibility – only interoperability of descriptive metadata. The OGC [127] defines interoperability as the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units (ISO [128] 2382-1). The OGC definition applies to metadata interoperability, when one replaces the word "data" in the definition with the word "metadata". An important aspect is that interoperable metadata can be used by computer systems, in contrast to metadata that is designed to be read and understood by a person.

Benefits of Interoperability

Having interoperable metadata allows multiple systems to work with the same set of data and metadata. It helps ensure metadata records associated with one resource can be accessed, accurately interpreted and subsequently used by a system or integrated with metadata records associated with other resources.
For example, interoperable metadata allows:

- tools, such as address books, and drawing systems, to easily and accurately import data.
- people to move geospatial datasets between various GIS [129] systems.
- a dataset to be searched and found through multiple catalogs without the provider having to implement multiple sets of metadata.

Different institutions and projects will inevitably develop "customized" metadata templates and files. To ensure interoperability, these customized files must conform to essential standards, and include effective labeling. Once equipped with the standard and an effective labeling system, a technical tool can be implemented to translate (though sometimes imperfectly) between these customized metadata files (crosswalks). The existence of a vocabulary [130], such as a dictionary [124] and/or thesauri [105] can also aid in the development of interoperable metadata.

How to cite this Guide

Neiswender, C., Montgomery, E. 2009. "Metadata Interoperability—What Is It, and Why Is It Important?." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdataintro/mdatainteroperability. Accessed June 23, 2014.

Machine Readability

Description of one of the highly desirable criterion of metadata—machine readability

Metadata [5] provide important information about a data resource. In theory, this information can be provided in many forms. For instance, the methods section of a scientific journal paper can be considered metadata.

However, a long text description is usually a human-readable only format, not one that a machine can efficiently parse and understand. Machine-readable descriptions have a consistent and known structure in which specific items of information are labeled and appropriately separated, allowing discovery by electronic systems. A variety of formats can provide appropriate demarcation and separation of metadata elements [117] and values [119], including tab-delimited or comma-delimited text, and Extensible Markup Language (xml [131]). Once a computer system is given the key to your metadata (a machine-compatible description of the format that is used), it can point users to your data via the metadata.

Here's an example of metadata from the computer science world:

Say that you wanted to describe a laptop computer that you have: a PC running the Windows Vista operating system, and which has Microsoft Word, Microsoft Excel, and Mozilla Firefox software installed.

A machine-readable version of this information is

"Laptop Computer", "Operating System", "Windows Vista"
"Laptop Computer", "Platform", "PC"
"Laptop Computer", "Software", "Microsoft Word"
"Laptop Computer", "Software", "Microsoft Excel"
"Laptop Computer", "Software", "Mozilla Firefox"

Notice, a resource (Laptop Computer) can have multiple elements (Operating System, Platform, Software), and an element can have multiple values (Microsoft Word, Microsoft Excel, Mozilla Firefox).

Here's an example of metadata from the marine science world:

Say there is a multibeam dataset containing information from several multibeam instruments (SeaBeam 2000, SeaBeam 2112, and EM120), and providing data in two formats (MB32 and MB57).

Here's what it could look like in a machine-readable format:

"Multibeam Data", "Sensor System", "SeaBeam 2000"
"Multibeam Data", "Sensor System", "SeaBeam 2112"
"Multibeam Data", "Sensor System", "EM120"
"Multibeam Data", "Format", "MB32"
"Multibeam Data", "Format", "MB57"

In this case, a resource can have multiple elements, but when the information is represented in a machine-readable format, each line represents exactly one resource-element-value combination. These examples both use comma separated text (with quotation marks to delimit text vs number fields), but, as mentioned before, there are multiple ways to format machine-readable text.

How to cite this Guide

Neiswender, C., Stocks, K. 2011. "Machine Readability." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdataintro/machinereadability. Accessed June 23, 2014.

Getting Started - How You Can Publish Your Metadata

Practical guidance about publishing your metadata

There is no one-size-fits-all approach to metadata [5]. This guide is an overview of the steps in the metadata creation process from initial planning through publication in a metadata registry or repository. Even if your ultimate goal is not to publish your metadata, the initial steps below will still be relevant for planning your in-house metadata.

You will need to develop a metadata template, that is, a structure to organize your metadata that is based on standards and has been converted into an electronic format. You will then need to populate that template with your metadata, and then publish it to the outlet of your choice. This can be divided into the following steps:

1. Establish collaborations

To prepare for publishing your metadata, collaborate with the science and technical professionals in your organization. Your metadata project will benefit from both types of expertise, so initiate the collaboration early.

2. Consider metadata in early project planning

A common problem in writing metadata is not having the information required to actually fill in the metadata. To prevent this pitfall, plan the metadata before a new project begins. Developing your plan before any type of data collection or instrument deployment has occurred will assure both thoroughness and relevance.

Your plan should include how the metadata will be created, updated, and disseminated. Recording metadata accurately as the project proceeds is much easier than organizing it after it has been collected. For example, the metadata creation component should include a protocol [132] for field metadata collection including media, methods, and information to be collected. Knowing that information to be collected might include instrument settings or environmental conditions will make sure those metadata are not overlooked.

3. Evaluate your data and how it will be used

Get to know your data (or get to know someone who knows your data). Keep track of the questions that arise as you become acquainted with the data. Consider how you will use the data and what you will need to know to use it.

By creating metadata, you help promote the long-term use of the data set you are describing. When planning a metadata approach, think about how future users might view your metadata.

For example, consider a hypothetical situation ten years from now when a researcher is conducting a study that requires the blending of multiple oceanographic data sets. To do the study, the researcher must verify that your data was collected using appropriate instrumentation and that appropriate techniques were used for the post-processing (for example, calibrations, screening, algorithms, and assumptions). Your metadata should be capable of supporting such a user. Understand your likely user group and prepare a long-term metadata strategy.

4. Select one (or more) metadata standards

Metadata standards are formal descriptions of the content and, in some cases, the format that metadata should have (See Metadata Standards for more information on finding and selecting metadata standards). Adopting an existing metadata standard can make your job easier by providing an established template and tools, which will make it easier to manage, share, and archive both your data and metadata. It is important to be aware of the standards that apply to metadata in general, and to your project in particular, and which standards are used in your domain. If a standard seems confusing, redundant, or contradictory, consult with experts to clarify inconsistencies before implementing it. See the guide on Selecting a Standard for more information.

5. Create a metadata template that fits your data

Your goal is to develop a list of metadata elements [117] that, when completed and associated with a particular data resource, will completely describe your data and put it in context with similar projects. Examples of common metadata elements include the names of the parameters measured, the location of the measurements (generally latitude, longitude and depth), and the contact information of the data provider. Use your scientific expertise to determine how best to describe your data using the selected standards as a general framework. Some of the elements included in a standard might be mandatory. Some of the optional elements might be appropriate, while others might not. Identify if there are any conflicts with the selected standards where the description of your data doesn't fit into the standard. You may wish to supplement the core elements from a standard with elements you think are critical to your specific data.

Ideally, your template will facilitate the creation of metadata that makes sense for your data, adequately allows for discovery [113] and reuse of the data, and appropriately satisfies reporting requirements. It is important to note that your set of metadata elements may evolve over time. While planning ahead is important, you might later realize that you actually need another element or that you have information that you don't need. It is fine to adjust as needed.

Tips for selecting metadata elements for your template

- Review the questions you asked when you were getting to know your data. Answers to those questions will most likely help the scientific community find and use your data, so this information should be included in your metadata.
- Look at what other projects have done. Search for their data and examine the metadata, making comparisons with your project.
- Review the documentation associated with the relevant standards. Some standards will include sample templates.

Helpful tools

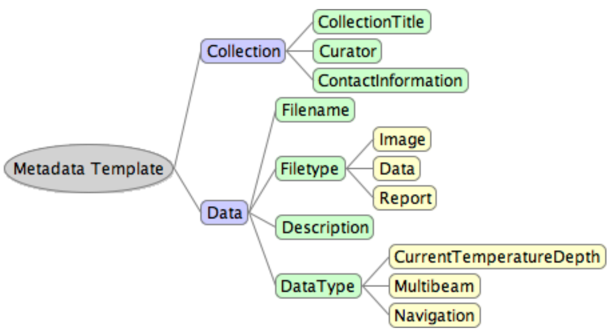
As you are exploring the standards, you will want to develop a working list of elements. Two useful tools for this are Microsoft Excel and

Freemind.

Excel [133] provides a tabular approach to your metadata template. The simple example template below is divided into two major metadata sections: Collection and Data. Your metadata will include more major sections. In this example, the first column provides the general category of metadata, the second provides the list of elements that will be included, and the third gives controlled vocabularies [122] where appropriate.

Collection	CollectionTitle	
	Curator	
	ContactInformation	
Data	Filename	
	Filetype	Image
		Data
		Report
	Description	
	DataType	CurrentTemperatureDepth
		Multibeam
		Navigation

Freemind [134] provides a graphical approach to your metadata template. The figure below shows the same metadata information in the table above, expressed as a Freemind map.



There are a variety of other tools developed for metadata template development. Some of the more widely used metadata standards have existing templates or entry interfaces to facilitate their use.

6. Determine your metadata format

Once the template includes metadata elements and the beginnings of controlled vocabularies, then it needs to be deployed into the project's technological infrastructure. In other words, how you capture the metadata digitally and link it to your data objects in a way that can be searched and used by humans and computers.

Storing your metadata

Choosing a format for storage of your metadata will affect long-term accessibility. One of the most versatile formats for your metadata is a comma separated value [119] (CSV [135]) flat text file (ASCII [136]). By storing your metadata in a CSV ASCII file, your metadata may be deployed and translated into a variety of formats. Using the metadata example above (see the Excel and Freemind diagrams), the corresponding CSV metadata file might look like this:

"Collection_CollectionTitle","RV Melville Cruise MV0909"
"Collection_Curator","SIO Geological Data Center"
"Collection_ContactInformation","http://someurl.edu"
"Data_Filename","NavFile.txt"
"Data_Filetype","Data"
"Data_Description","This is sample metadata for text-formatted navigation data"
"Data_DataType","Navigation"

In this example system, you would generate a single text-metadata file for each data document in the system.

Because delimited text is harder for the human eye to scan and work with, entering data in an Excel spreadsheet sheet and exporting it as delimited text is useful. Extensible Markup Language (XML [131]) is another text-based format that many standards are adopting.

Ideally, you want to keep your metadata closely coupled, or linked, to your data object so that the metadata aren't lost when the data are moved or distributed. Headers within files, zipped or tarred bundles that include data and metadata files, relational and other databases, and Excel workbooks with sheets for data and sheets for metadata are all approaches for doing this. If you are sharing your data with others, or archiving it for long-term preservation, it is better to avoid proprietary formats like Excel and relational database management systems in favor of text formats. Your decision will be based on the nature of your data (some data are more suited to flat files, others are more relational) and the infrastructure capabilities of your project or institution.

7. Implementation - capturing your metadata

Once you know how you will record your metadata, you will need to develop a plan to capture it. For each piece of metadata, consider how it will be recorded, by whom, and when. In some cases, you can draw metadata (and data) directly from instruments. In other cases, a researcher will have to note down information. Determine the most efficient process, test it on a small scale, make adjustments as needed, and then implement it to gather your metadata.

8. Check your metadata

Once your metadata has been entered into your system, it should be reviewed for consistency, completeness, ambiguity of terms, appropriateness of title, and readability of input. See the guide on Writing Good Metadata for more information.

9. Publish

The previous steps were preparatory to the final goal of sharing your data with your research community through publication. A typical location to publish your project metadata would be a metadata clearinghouse [137], also called a registry or repository. The National Ocean Data Center and the Global Change Master Directory are two examples of clearinghouses. Clearinghouses permit automatic searching of metadata, so that researchers can find out that your data exist. Ideally, the metadata should be available to the international community. This will provide maximum exposure of the quality data sets that you and your organization produce. If you choose not to use a clearinghouse, you can make your metadata available to your target audience through some other means, such as a web-based portal, though we strongly recommend registration in addition to individual website creation.

After you have completed the uploading process to your location of choice, check to make sure the publication process was a success. Also, check the search capabilities of the clearinghouse to confirm that your entry will be visible or discoverable in ways that you expect. If it isn't and you don't understand why, contact the clearinghouse for help and clarification.

10. Participate in the MMI community as an experienced metadata publisher

Share your success with the marine metadata community. Although this is not part of the publication process, participating in the larger community and communicating your experiences will help improve metadata processes. You can contribute a case study about how developed your metadata and what you learned in that process, or suggest additions or changes to these guides.

How to cite this Guide

Neiswender, C., Isenor, A., Stocks, K. 2010. "Getting Started - How You Can Publish Your Metadata." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdataintro/gettingstarted. Accessed June 23, 2014.

Meeting NSF Data Policy Requirements

How to conform to the National Science Foundation's Division of Ocean Sciences data sharing policies.

The National Science Foundation [138] (NSF) provides a brief Division of Ocean Sciences Data and Sample Policy [139] (2004) outline. This policy applies specifically to the NSF's Division of Ocean Sciences. Other NSF program policies may differ.

All Principle Investigators (PIs) receiving NSF Division of Ocean Sciences grant money are required to share their data. This should be done as soon as possible, but **must** be done no later than two years after the data is collected. In addition, as of October 2011, NSF has a new requirement that all proposals must include a two-page data management plan [140].

Data must be submitted to the appropriate National Data Center [141] relevant to your area of research (see list below), and data are to be submitted according to formats and via the media designated by each national center. If no appropriate data center exists, principal investigators must develop an alternative means to share their data.

National Oceanographic Data Center (NODC [142])

- Data submission instructions [143]
- Long version of data submission guidelines [144]

National Climatic Data Center (NCDC [145])

- NOAA Paleoclimatology data contribution [146]
- Contributing dendrochronological data [147]
- Contributing tree-ring fire event data [148]

National Geophysical Data Center (NGDC [149])

- NOAA Paleoclimatology data contribution [146]
- IHO Data Center for Digital Bathymetry data submission [150]

National Snow & Ice Data Center (NSIDC [151])

- Data Submission [152]

Carbon Dioxide Information Analysis Center (CDIAC [153])

- Data Submission [154]
- It is recommended to review this Best Practices [155] document and then contact one of the the CDIAC staff members [156] to discuss further data submission plans.

How to cite this Guide

Gale, A., Stocks, K. 2010. "Meeting NSF Data Policy Requirements." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdataintro/meetingnsfdatareqs. Accessed June 23, 2014.

Writing Good Metadata

Checklist for metadata creation

To create quality metadata [5], it is essential to include all the descriptive information necessary to locate, understand, and use a data set. You can save time and resources by leveraging the work and tools developed by others. Write your metadata in a standard [157] fashion and adhere to the following rules and considerations (based on excerpts from the NOAA Coastal Services Metadata Workshop Materials [159]).

- **Write simply, completely, and consistently**. This will help general audiences understand your metadata and, therefore, help them understand your data set. This will also enhance the automatic search capabilities, which in turn will attract a wider audience for your data. You need to use the same labels (names) for your metadata elements [117] every time they are used, and the same format for your metadata (such as comma delimited text). Your metadata descriptions need to be internally consistent and, ideally, consistent with descriptions produced by other organizations in a similar discipline.
- **Use controlled vocabularies.** Controlled vocabularies help limit the terms. For example, consider data coming from research vessels. If one participant on the cruise enters the vessel name as RV Revelle, another as Roger Revelle, and another as R/V Revelle, there will be no way for someone searching the metadata later to easily find all the data coming from this vessel. A controlled vocabulary that specifies the format of the vessel name, as well as other terms that can be predicted in advance, will ensure standardization. There are many recognized vocabularies [130] within the marine domain, and by using terms from these vocabularies, you will create metadata that is consistent not just within your dataset, but across other datasets.
- **Provide an appropriate descriptive title for your dataset**. A descriptive title of the contents is important for those trying to decide whether or not to explore the actual data set. You might consider a title that summarizes important features of your data, answering “what, where, when, and who” questions about your data. Scale, location, and date are important factors to consider in your title information, e.g., “California bathymetry data 0 –200 m, 10m contours, 2004.”
- **Clearly state data limitations**. Limitations of your data set are very important for establishing the relevance and usage potential of your data set. A common example is “not to be used for navigation purposes.”
- **Choose unambiguous, descriptive keywords**. Some metadata standards have a specific element to hold keywords—terms on which a user can search in a metadata registry or repository. They are essential for users trying to locate your data, so they should be chosen with care and be drawn from suggested keyword lists when available.
- **Avoid using special characters.** Don’t use printing or non-printing characters that might be misinterpreted by a computer. Printing characters include !, @, <, >, (,), while non-printing characters include tabs and carriage returns. This will help make your metadata machine-readable [112].
- **Review your metadata for accuracy and completeness**. Have a second look at your metadata and have someone else take a look at it as well. When reviewing your metadata, ask the following questions:
  - Could someone use an automatic search to locate this data set?
  - Could they assess its usefulness?
  - Do your metadata include enough specific information to uniquely identify and locate any geospatial data based solely on your documentation?
  - Can a novice understand what you wrote?
  - Does the documentation adequately present all the information needed to use or reuse the data represented?
  - Are any pieces of information missing, such as projection information, source citations, and process steps?
  - Are your key words descriptive enough to help other people find your data set?
  - Have you used enough broad terms? Have you used enough narrow terms?

How to cite this Guide

Isenor, A., Bermudez, L., Watson, S. 2010. "Writing Good Metadata." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdataintro/writegoodmdata. Accessed June 23, 2014.

Some Examples of Metadata

Sample metadata records, in a variety of formats

Metadata [5] can apply to just about any data or information about your project; for instance [160], sensor configuration, data from the sensors, images, publications, or GIS [129] layers. Therefore, it is helpful to have some examples to help plan what should go into your metadata records. The links below provide examples of metadata in several commonly used conventions from a variety of marine-related fields. Some pages provide links so the user may view the XML [131] or text versions instead of the default HTML [162]-rendered view.

- FGDC Metadata describing the South Florida Information Access (SOFIA) website [163]
- FGDC Metadata a dataset of hydrology and water quality in the Everglades, presented as part of the SOFIA site [164]
- FGDC Metadata describing the NOAA collection of 30 meter Digital Elevation Model data of estuarine Bathymetry [165]
- FGDC description of Sea Surface Temperature data from MOOS Upper-Water-Column Science Experiment (MUSE) from NOAA-14 AVHRR Satellite [166]
- FGDC metadata on CTD , Optical backscatter, Fluorescence, Bioluminescence data from MBARI MUSE AUV Data [167]
- Harmful Algal Bloom Mapping System (HABMapS) and Sea Surface Temperature (SST) Data derived from satellite in the Gulf of Mexico [168]
- Chesapeake Bay Program Water Quality Database [169]
- DIF metadata example from the Global Change Master Directory (GCMD) portal [170]
- Long Island sound Benthic Formainifera distribution (ARCVIEW SHAPEFILE) [171]
- USGS InfoBank display of FGDC metadata describing the CTD Data from a joint NOAA/USGS research cuise in the Caribean Sea [172]
- MERSEA display of ISO19115 metadata for hydrographic data(the XML version may be viewed by clicking on the button at the top left of the page) [173]

More examples are available in the metadata content standards [174] section of our site.

How to cite this Guide

Montgomery, E., Bermudez, L. 2010. "Some Examples of Metadata." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdataintro/mdataexamples. Accessed June 23, 2014.

Metadata Standards

Introduction to content and format standards

A metadata [5] standard [157] is a model for metadata storage that is approved by a recognized standards organization, such as ISO [128] or FGDC [111]. Metadata standards specify the kinds of information required to describe data. When a metadata document conforms to a standard, it is considered formal metadata [175]. Standards can provide very specific information about details such as values [119] to be provided and how to technically present the metadata.

Metadata standards can be of either or both of these two general types:

- Content standards [176] (also called descriptive standards)
- Format standards [178] (also called technical standards)

Content Standard

A content standard specifies the information required to document a data set. It is a list or hierarchy of required metadata elements [117] to be included in the metadata description, where the metadata element is considered the basic unit. Content standards include the name of each metadata element (also called the label [180]) and a definition for each name. For example, a common oceanographic metadata element is vessel. The definition would specify that this element should hold the name of the research vessel used to collect the data, and might further specify that the name be drawn from the ICES list of ship names. Each metadata element contains specific information, which when combined with content from other elements appropriately describe a data set.

The content standards also requires a set of statements that formally express the rules of usage for the collection of metadata elements. For example, a rule might specify that a particular element is mandatory, or that if one element is included (such as a parameter name) that another element must also be provided (such as the units for the parameter). approved model—the definition of metadata elements and their rules for use—when implemented by multiple projects, helps ensure common practices across existing projects and helps users develop a sound metadata plan for a new project.

Format Standard

While content standards describe the information that should be captured, format standards express how that information should be represented. HTML [162], XML [131], NetCDF [181], and delimited text are all common formats. Format standards are critical for allowing machine readability [112] of metadata. Some metadata standards, such as ISO 19139, are both format and content standards.

Use of Content and Format Standards

The integration of the content standard and format standard concepts allows for consistency of information that is included in a metadata set, the structure that is used to store the metadata, and where the metadata are stored within this structure. By carefully articulating how the metadata elements are named, structured, and utilized, metadata standards enable interoperability [7] and allow for the creation of tools to work with the metadata, such as searchable repositories and metadata creation templates. See more about this topic in The Importance of Metadata Standards.

Additional Reading

For an interesting discussion of formal metadata, see the Metadata Challenges presentation by Dr. Sara Graves.

How to cite this Guide

Neiswender, C., Isenor, A., Graybeal, J., Hogrefe, K., Bermudez, L., Galbraith, N., Watson, S. 2011. "Metadata Standards." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards. Accessed June 23, 2014.

The Importance of Metadata Standards

Discussion of the merit of adherence to metadata standards

Metadata [5] Standards are sets of topic-specific norms and definitions that guide the collection and documentation of metadata so that the result has consistent collection criteria, nomenclature, and structure. This consistency defines formal metadata [175]. Creation of formal metadata through adherence to standards is essential for sharing data, for data management within the project, and for future funding.

Formal metadata standards enable interoperability [7] among similarly formatted databases on local and global scales, greatly enhancing the reach of scientific research through data sharing. In order for data to be fully interoperable, consistent ontologies [107] must be used throughout data sets—metadata should use the same terms for describing a particular type of scientific inquiry, natural phenomenon, or analytical method. Data sets that do not have consistent terminology between them may still be made interoperable by using ontologies to associate terms with common definitions. (For more information, please see the Ontologies section.)

Interoperability also involves standards of database structure and organization. The combination of vocabulary [130] and database standards enables reliable queries during data searches and the use of common analytical models, allowing for collaboration between projects and individuals that are studying similar concerns.

In addition to collaborative benefits, metadata standards also provide specific advantages to the individual project. A good metadata standard results from the broad consultation of experienced researchers in a particular field [117]. When combined with robust scientific design, field protocols [132], and data processing methods, the standard ensures quality data collection and documentation.

An established standard provides the definition of the geographic, environmental, and equipment-related information that should be recorded while in the field. Standards also provide direction for the documentation of data processing, file naming conventions and formats, and a glossary [13] of precise definitions for applicable terms. These provisions allow the researcher to maximize the use of tools designed to facilitate data interoperability through quality data collection and organization.

Data collected in accordance with the quality and organizational guidelines set out in an established metadata standard are not only more easily shared, the studies that rely on such data are more easily funded. An increasingly common way to share and seek data on a particular topic is through the use of data clearinghouses or geo-portals that serve as cyber-depots for the sharing of data and collaboration between scientists of related interests.

The consistency that is provided by the use of standards is essential for such sites to function, and adherence to a standard is most likely a prerequisite for participation. Additionally, many funding agencies are now requiring the use of metadata standards in general, or a particular standard, in projects that they support to assure that the data from studies that they fund can be easily integrated into existing databases and GIS [129] projects.

How to cite this Guide

Hogrefe, K., Stocks, K. 2011. "The Importance of Metadata Standards." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/stdimportance. Accessed June 23, 2014.

Metadata Standards vs. Metadata Specifications

Distinguishing between a metadata standard and specification

Some confusion exists about the use of the words *standard* [157] and *specification* when applied to metadata [5]. On the MMI [9] site, we maintain the following distinction:

A *metadata standard* is a set of rules that define the creation of metadata. These rules are formally usually expressed in a document that outlines acceptable usage of the elements described within the standard. Documentation of these elements typically includes things such as name, definition, and structure. These rules also have formal approval, publishing, and governance procedures, as established by a formal body or organization with broad community-based representation.

A *metadata specification* is any description of how to store metadata. A specification can be developed and implemented by any level of an organization. A specification does not need formal documentation, nor does it need broad community-based approval. There may or may not be a formal governance procedure for a specification. Governance is not required. If it does have governance, it is often at a local or even an individual level.

Specifically, all standards are specifications [104], but not all specifications are standards. Because a standard is governed by a large community of stakeholders and is well documented, published, and governed, it promotes interoperability [7] between organizations that use the standard and, therefore, provides greater value [119] to the international community.

How to cite this Guide

Isenor, A., Neiswender, C., Graybeal, J., Stocks, K. 2011. "Metadata Standards vs. Metadata Specifications." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/stdvsspec. Accessed June 23, 2014.

Common Metadata Standards

Widely used, broadly applicable content standards

This guide describes several standards that offer a good starting point for those seeking a generic metadata [5] approach to their data: ISO [128] 19115 and ISO 19139, CSDGM [182], DIF [183], Dublin Core, and ADN [184]. This is not an exhaustive list, but it presents a selection of standards that have the following characteristics:

- *Widely-used*: adopted by major national or international organizations or communities of practice
- *Broadly-applicable*: designed to describe a variety of different types of data that marine users might have
- *Maintained*: actively supported by a standards body and extended and updated as appropriate
- *Applicable to content*: content standards [176] that describe the information that must or can be ascribed to a data set. (This is in contrast to format standards [178] and transport protocols [132].)

ISO 19115 and 19139

Developed by the International Organization for Standardization, ISO 19115 is designed to describe geographic information and services, including extent, quality, spatial and temporal schema [185], spatial reference, distribution, and identification information. It contains both mandatory and optional components that are organized into sections, and it has defined methods for extending the standard to fit specialized needs. The full description of ISO 19115, like all ISO standards, may be purchased from the ISO website. See MMI [9]'s ISO 19115 page for more information. It is a content standard only; a related standard, ISO 19139, defines a format standard called Geographical Metadata XML [131] for holding ISO 19115-compliant content. ISO 19115 and 19139 are widely used.

North American Profile of ISO 19115

This standard was developed to replace the CSDGM standard (see below) as the US federal metadata standard for geospatial data. As a profile based on ISO 19115, it is applicable to any kind of geospatial data. Metadata that are compliant with NAP are fully ISO 19115 compliant as well. For links to further information, including the current status of NAP with respect to federal approval, see the MMI reference for NAP.

CSDGM (Content Standard for Digital Geospatial Metadata)

The Content Standard for Digital Geospatial Metadata, often informally called the FGDC [111] Metadata Standard, was developed and is maintained by the Federal Geographic Data Committee, a US governmental organization. It is the official US Federal metadata standard and is mandatory for geospatial data produced by US Federal agencies (though the North American Profile of ISO 19115, presented above, is under consideration as a replacement standard). It has also been adopted by many US state and local governments. Like ISO 19115 and 19139, it is applicable to any geospatial data. See MMI's CSDGM page for more information.

DIF (Directory Interchange Format)

One of the oldest standards, DIF began its development in an Earth Science and Applications Data Systems Workshop held in 1987 as a means of creating catalog interoperability [7]. The International Directory Network, the body behind the CEOS IDN Master Directory, now maintains DIF and uses it as the metadata standard for its directory. DIF is one of the smaller standards (with respect to the number of elements), having 8 mandatory elements and 36 total elements. The mandatory elements are EntryID, Entry Title, Keywords, ISO Topic Category, Data Center, Summary, Metadata Name, and Metadata Version. The Keywords element is supported by a hierarchical, controlled vocabulary for categorizing earth and atmospheric sciences resources. DIF is designed to work with several other standards. It is compatible with both ISO 19115 and FGDC's CSDGM. See MMI's DIF page for more information.

Dublin Core

Another of the smaller standards, the Dublin Core was originally developed as a way to describe bibliographic data, but now is used more widely, as it provides substantial content flexibility. It has two forms: the simple Dublin Core and the qualified Dublin Core. The former was developed first and consists of 15 standard elements: Contributor, Coverage, Creator, Date, Description, Format, Identifier, Language, Publisher, Relation, Rights, Source, Subject, Title, and Type. All of the elements are optional, and there are no restrictions on how the elements are filled in.

The qualified Dublin Core has three additional elements: Audience, Provenance [108]

The record of how a particular value or record came to be. Provenance can include things like when, by whom, and how the item was created or modified.

, and RightsHolder, as well as a group of qualifiers that either define the meaning of elements more narrowly and specifically or provide encoding schemes. Encoding schemes have further constraints, such as:

- Requiring that entries for an element be taken from a particular controlled vocabulary; for example, requiring that subject terms be taken from the Library of Congress Subject Headings [186], or another of the approved subject lists.
- Specifying a format; for example, requiring that dates be entered as yyyy-mm-dd.

ISO 15836 is the ISO implementation of the Dublin Core standard. While Dublin Core is a content standard only, the Dublin Core Metadata Initiative, which maintains the standard, also provides encoding guidelines for RDF [187], XML, and HTML [162] format standards in support of their content standard. See MMI's Dublin Core page for more information.

ADN

ADN is the ADEPT/DLESE [188]/NASA metadata framework. It was originally developed to describe educational resources, such as lesson plans and classroom activities, for discovery [113] by earth sciences educators. It is now being applied to a much wider variety of data, and emphasizes geospatial and temporal aspects of data. It includes a small set of required elements that the metadata cataloguer must include, such as title, subject, and resource type; another small set for the resource creator to provide, such as language and terms of use; plus a larger set of optional elements that can further describe the content. Many of the elements have controlled vocabularies [122]. See MMI's ADN page for more information.

How to cite this Guide

Neiswender, C., Isenor, A. 2011. "Common Metadata Standards." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/comstds. Accessed June 23, 2014.

Standards Bodies

Coming soon... Watch for updates!

One of the primary questions to ask when considering a metadata [5] standard [157] is who developed the standard. A standard that was developed with broad, organized community input, and which has an organization to manage and maintain it is more likely to be useful, widely adopted, and persistent into the future. Here, three of the larger standards bodies are describe: the International Standards Organization (ISO [128]), the Dublin Core Metadata Initiative (DCMI [190]), and the Federal Geographic Data Committee (FGDC [111]).

ISO [191]: International Standards Organization

Focus: Content and Format Standards

The world leader in the effort to set international standards is ISO [192], which works toward consensus agreements between national delegations representing all the economic stakeholders concerned—suppliers, users, government regulators and other interest groups, such as consumers. They agree on specifications [104] and criteria to be applied consistently in the classification of materials, in the manufacture and supply of products, in testing and analysis, in terminology and in the provision of services.

ISO provides a reference framework, or a common technological language, between suppliers and their customers, which facilitates trade and the transfer of technology. The American National Standards Institute (ANSI [193]) is the US representative of ISO.

The ISO standards most relevant to marine data are

ISO15836	The official adoption of the DCMI core metadata element set as an ISO standard Technical Committee 46 (Technical Interoperability [7])
ISO19115	Geographic information metadata content and conceptual framework Technical Committee 211 (Geographic Information, Geomatics)
ISO19139	Geographic MetaData (gmd) - XML [131] schema implementation Technical Committee 211 (Geographic Information, Geomatics)

DCMI: Dublin Core Metadata Initiative

Focus: Largely Content Standards

DCMI represents a focused effort to establish a baseline for metadata content standards. DCMI takes its name from its origins as an invitational workshop convened in Dublin, Ohio, in 1995 with the purpose of establishing “core” elements that should be included in any metadata set because they are broad, generic and useful for describing a wide range of resources. The “Dublin Core” consists of 15 basic elements that are considered part of a larger metadata resource maintained by DCMI.

In addition to the core elements, DCMI metadata terms include a broader set of elements and element refinements, a vocabulary [130], and a list of encoding schemes (Dewey Decimal, Library of Congress, etc.). They also provide for encoding guidelines for employing RDF [187], XML and HTML format standards (schemas [185]) in support of their content standards.

It is intended that these resources allow the combination of DCMI standards with other compatible standards in the creation of application profiles [115] based on the DCMI Abstract Model. DCMI also serves as an active promoter of support services and fostering active forums to provide examples of successful applications of the core standards in various fields.

Major Standards

ISO15836	The official adoption of the DCMI core metadata element set as an ISO standard Technical Committee 46 (Technical Interoperability)
ANSI/NISO Z39.85-2007 [194]	Relies on DCMI Core metadata element set to define fifteen metadata elements [117] for resource description in a cross-disciplinary information environment.

FGDC [195]: Federal Geographic Data Committee

Focus: Largely Content Standards

The focus of the Federal Geographic Data Committee is to “provide coordination of federal geospatial activities between, among, and within agencies by establishing policy and providing guidance and direction” to United States governmental agencies.

The FDGC developed and endorsed the Content Standard for Digital Geospatial Metadata (CSDGM [182]) in 1994 and updated it in 1998. At the time this guide was written, all federal agencies are required to use this standard in metadata documentation. Many state and local agencies have adopted it as well to assure "up the line" compatibility and to comply with reporting requirements tied to funding.

In order to ensure international interoperability, FDGC is in the process of adopting the North American Profile of ISO 19115 as its primary standard.

FGDC also endorses a suite of profiles and extensions to CSDGM, including an extension for remote sensing data, a biological data profile, and a shoreline data profile. These allow the creation of metadata tailored to the specific characteristics of these data types, while still providing full compatibility with the base standard.

How to cite this Guide

Hogrefe, K. 2011. "Standards Bodies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/comstds/standardsbodies. Accessed June 23, 2014.

Discipline-Specific Standards for the Marine Community

The following describes some standards that have been created for specific topics relevant to the marine community. We use the term standard here in its more general sense to include full standards, as well as extensions and profiles [115].

Marine Community Profile of ISO 19115

This profile was developed by the Joint Facility Australian Oceanographic Data Centre and covers the topic of oceanographic data sets related to sea-going collections. It was proposed (draft version 1.3 August 2007) as the basis for an International Oceanographic Data Exchange (IODE) Marine Community Profile. If accepted, this profile would become the metadata [5] profile used by the international oceanographic community working under the auspices of the Intergovernmental Oceanographic Commission (IOC) of the United Nations.

The profile defines several extensions to ISO 19115 that are divided into three categories:

1. The addition of metadata elements [117] that describe the temporal aspects of both the metadata and the data. This includes when the metadata record was last revised, how frequently a sample was taken, whether the data are aggregated over a specified time span such as weekly or monthly, and whether the dataset in question is the most up-to-date version.
2. Changes in element conditions to allow the profile to specifically address traditional procedures for oceanographic data collection. This includes making an element that is optional in the core standard mandatory in the profile. For example, oceanographic data are often identified using some type of unique numbering system. The profile makes the fileIdentifier element mandatory rather than optional, thus uniquely identifying the metadata record. This is critical when large numbers of metadata descriptions are distributed among interested parties, as it allows the parties to identify unique or non-unique descriptions. The geographicElement and temporalElement have also been made mandatory if the resource being described is a data set. This change makes mandatory the description of the temporal and geographic extent of the data set, thus directly associating the data set with a traditional sea-going collection.
3. Changes in code lists [196] to help the metadata author describe the temporal currency of the data set. In ISO 19115 terminology, a code list is a controlled vocabulary [122] for the content of a particular metadata element. In this profile, the currency code list defines "most recent," "historical," and "predicted" as types of temporal currency. A code list is also defined for the temporal aggregation of the data with values [119] such as "day," "multi-day," "week," "month," etc.

FGDC CSDGM Profile for shoreline data

The FGDC CSDGM has been used by the coastal shoreline and biological communities in the US to define an application-specific profile that clarifies important aspects of shoreline data. These important specifics result in modifications to the CSDGM standard related to both occurrence and domain.

In many cases, the metadata elements related to time have been changed from "mandatory if applicable" or "optional" to "mandatory." This change highlights time as a critical component in the collection of shoreline data.

Other domain changes continue to emphasize the importance of time. The restriction of some metadata element content to be specific to local time is intended to increase consistency across metadata descriptions. Domain changes were also made in the area of horizontal and vertical positional accuracy reports and explanations. These changes specifically relate the explanatory text to the US National Standard for Spatial Data Accuracy (NSSDA). This relationship [125] is intended to encourage or enforce compliance with the positional accuracy steps outlined in the NSSDA.

Metadata elements were also added. Again, these specifics are directed at important aspects of the conditions that could impact the resulting data set. The added elements include information on tidal conditions, weather conditions, and environmental events. Tidal information data include type and time of the tide, while the weather condition elements include wind speed and direction, wave height, and barometric pressure.

The environmental event descriptor is used to document substantial events (for example, hurricanes) that have recently taken place in or near the data collection site. All of these factors are important metadata descriptors that help potential users assess the data set for their particular application.

How to cite this Guide

Isenor, A. 2011. "Discipline-Specific Standards for the Marine Community." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/comstds/disciplinespecificprofandext. Accessed June 23, 2014.

Metadata Extensions and Profiles

Metadata Extensions and Profiles

Customizing metadata standards with extensions and profiles

The Common Standards guide describes widely used standards for projects seeking a generic metadata approach. These standards were generally created for broad application, and thus usually represent the lowest common set of metadata that applies to many disciplines. These standards provide an organized way to represent many of the general characteristics of a dataset.

However, a general standard cannot include every element or relationship [125] that is important to every discipline within a broad subject area. To accommodate specialized needs, users can either select a more specific metadata standard (one developed for a specific kind of data or use), or a metadata extension or profile that is based on a more general standard.

Special metadata elements [117] that contain discipline-specific information can be added in ways that comply with the rules of the more general standard. These additions are called extensions and result in the creation of a profile for a specific discipline to which the standard is then applied.

Metadata standards are typically managed in a way that supports the creation of specialized forms of the standard to meet the needs of a particular community. However, the changes must not make the standard less strict. Changes may not alter the particulars of the original standard but provide a definition for an extension or a profile of that standard.

To demonstrate an extension making a standard more strict, suppose a metadata standard has an optional metadata element named sensor_owner. It Because it is optional, a metadata creator may or may include it in project metadata. However, one particular organization that creates metadata using this standard, considers this information to be critical. Thus, the organization mandates that their metadata records will always contain the sensor_owner element. This mandated field is considered an extension to the original standard, and the specialized form of the standard created by the addition of the extension is called a profile. The organization has extended the metadata standard to meet their particular needs, while not contradicting any required content as specified by the standard.

Standards can be extended in other ways. Sometimes a metadata standard does not provide all the elements required by a community. In this case, the community can extend the standard by adding new elements.

For example, if a project were collecting shoreline data, wave height might be an important piece of information to convey. Wave height is relevant in a few specific environments but not for all environments, so a general standard might not include an element for wave height. One solution might be to include wave height in an existing free-text element within the more broadly-applicable standard. For example, wave height could be included within the Description element of FGDC [111]'s Content Standard for Geospatial Metadata or in the Summary element of the Directory Interchange Format. The problem with this approach is that unstructured information (that is, free text) can't be easily searched or categorized by a computer. So, if wave height is an important enough detail, a it can be added as an extension, as long as it complies with the rules of the standard.

In all cases it is important to keep in mind that community-based extensions must follow strict rules for extensions as set by the original standard.

Types of Extensions

A metadata extension is the sum of additions to a metadata standard that allow users to provide information in additional fields or in additional ways that were not specified in the original standard.

In standards such as ISO [128] 19115, extensions include the following types of changes:

- Addition of a new metadata section
- Alteration of the domain of a metadata element (for example, assigning a code list [196] to specify what responses are allowed for that metadata element)
- Addition of terms in a code list
- Addition of a new metadata element to an existing metadata element
- Addition of a new metadata entity
- Changing the obligation of a metadata element from optional to mandatory (but not the reverse, which would break the core standard)

Constraints are considered a specialized subset of extensions, in which additional restrictions are placed on the standard. In the above list the second and last items are constraints. In this case the term “extension” describes the addition of information to the standard, even though the metadata instances that follow the standard are thereby restricted. More succinctly, the standard is expanded, but the metadata is restricted.

Profiles

A metadata profile is the community-specific application of a metadata standard. Profiles must meet the core requirements of the metadata content standard; that is, they must provide the mandatory elements that the standard requires. But they also include extensions (described above).

A metadata content standard is composed of the core metadata set and optional elements:

```
metadata content standard = core metadata set + optional elements
```

Therefore, a profile also can be thought of as the sum of these component parts:

```
profile = core metadata set + optional elements + extensions
```

Profiles are used to adopt a more general standard to a particular domain-specific purpose. One example is the Shoreline Profile for CSDGM [182]. The CSDGM metadata standard is designed to document a wide variety of geospatial data. The Shoreline Profile was created to better tailor CSDGM to data that were taken from a shoreline location or from a location that intersects a shoreline. This profile includes the following changes: changing some elements that are particularly important for shoreline data from optional (in CSDGM) to mandatory; requiring that certain keywords be taken from an approved shoreline glossary [13] instead of allowing free text; and adding new metadata elements, such as Time_of_Tide and Wave_Height, that can be relevant to shoreline data, but are not relevant to all geospatial data.

These examples of metadata profiles demonstrate how profiles relate to the original standard:

- Biological Profile for CSDGM [197]
- World Meteorological Organization profile to ISO 19115 [198]
- Marine Community Profile to ISO 19115 [199]

See the MMI [9] Guide on Discipline-specific Profiles and Extensions for the Marine Community for more information on this topic.

Management of Extensions and Profiles

In most standards organizations, a formal metadata [175] extension or profile must be approved by the organization that approved the original standard. This complex process can take some time, but it results in official approbation" that will reassure users about the validity of the extension or profile. Also, this process provides a way for the approving organization to publicize the standard, along with any extensions or profiles, in a single web site or reference.

Some groups might want to extend a standard to meet their particular needs without going through the lengthy review and approval process. This option is feasible, but there are disadvantages to this approach:

- Communities won't be able to locate the extensions and profiles as readily
- Profiles with additional elements can break software that doesn't know about those elements
- Some metadata standards may have legal restrictions on their re-use in a non-approved profile or extension

Value of Extensions and Profiles

The developers of most content standards [176] expect and encourage the development of extensions and profiles and may direct how they are to be specified or registered. A community that adopts a profile increases the interoperability [7] of its metadata between groups within the community. Potentially, it even increases its interoperability with communities that use other profiles, because the use of the core metadata elements is shared. This view on interoperability is not shared by all members of the metadata community; some individuals feel that a proliferation of profiles and extensions reduces interoperability, since each group's metadata are not fully interoperable with metadata of other groups. These viewpoints must always be considered when deciding whether to confine metadata elements to those in the core standard, to lobby for changes to the core standard, or to customize your metadata with extensions and profiles.

How to cite this Guide

Graybeal, J., Isenor, A., Neiswender, C., Stocks, K., Bermudez, L. 2011. "Metadata Extensions and Profiles." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/extensionsprofiles. Accessed June 23, 2014.

Moving Between Standards (Crosswalking)

Introduction to crosswalks, including definitions and an outline of the crosswalk process

Crosswalks are human- or computer-readable documents that map metadata elements [117] between different metadata [5] standards.

Crosswalks can apply to content standards [176], vocabularies [130], or both. An automated crosswalk process may take an instance [160] of a metadata description that is presented in a particular format and change the format and element names and the values [119] within those elements (i.e., the vocabulary) to meet the requirements of the second standard.

Crosswalking is generally done when datasets using different metadata standards or vocabularies need to be integrated. For example, consider a website providing a searchable metadata directory. If the different datasets composing the directory were described using different standards and vocabularies, it would be difficult for a user to search across them effectively.

If someone was interested in wave height data, she might need to know to search for “wave ht (m)” in one dataset and “wave amplitude” in another. A crosswalk that defined these two elements as synonymous would allow a website to be constructed that allowed the user to search on either term, and retrieve applicable results from both datasets.

Due to the complexity of metadata content standards, there are few automated processes to crosswalk between content standards. Even in those cases where automated crosswalks exist, inevitably some information is lost when crosswalks are made. This is due to the complexity of the standards and potentially non-overlapping subject areas. When there are subject areas that do not overlap, even manual translation between standards does not result in complete information transfer.

The Crosswalk Process

The process of mapping between content standards or vocabularies is usually divided into the following steps: harmonization [200], semantic mappings [201], rules [202], and transformation [203]. These are described in the next guide.

Suggested Citation

Graybeal, J. [204], Isenor, A. [205], Wang, H. [206] 2009. "Moving Between Standards (Crosswalking)." In *The MMI* [9] *Guides: Navigating the World of Marine Metadata.* http://marinemetadata.org/guides/mdatastandards/crosswalks [207]. Accessed: 06/29/2010

How to cite this Guide

Wang, H., Isenor, A., Graybeal, J. 2011. "Moving Between Standards (Crosswalking)." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/crosswalks. Accessed June 23, 2014.

Harmonization of Metadata Standards

Description of harmonization

Metadata [5] standards are often described in terms of element [117] names and definitions. A standard defines the rules for how the metadata are structured and also the appropriate content for the various elements.

However, different standards can be stated in different ways. In other words, a particular standard (the source standard) doesn't have to use the same element labels (names) for similar content, or allow the same terms to be filled in to each element as another standard (the target standard).

In the harmonization [208] process, the source and target metadata standards are resolved with the same syntax or model. In the simplest case, this is done by creating a table of fields from each standard in a common application (e.g., a spreadsheet). The table rows would likely contain elements from the source standard that are in some way related to elements of the target standard. In the simplest case, there would be one-to-one relationships [125] between source elements and target elements.

In more complex harmonization cases, there are one-to-many or many-to-one relationships. Also, intra-relationships between the elements within a single standard must be thoroughly described as part of the harmonization process. Of course, this implies the elements must be thoroughly described in the source and target standard.

How to cite this Guide

Wang, H., Isenor, A., Graybeal, J. 2011. "Harmonization of Metadata Standards." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/crosswalks/harmonization. Accessed June 23, 2014.

Semantic Mappings

Description of mapping relationships utilized in crosswalks

The term *semantic mapping* as applied to metadata [5] is a visual or tabular strategy for establishing the relationships [125] of vocabulary terms [209] between data sets.

Basic Relationships

When creating mappings among vocabulary [130] terms, the mapping organization requires a good set of basic relationships. The most common relationship, “is the same as,” is usually too narrow to adequately map all terms.

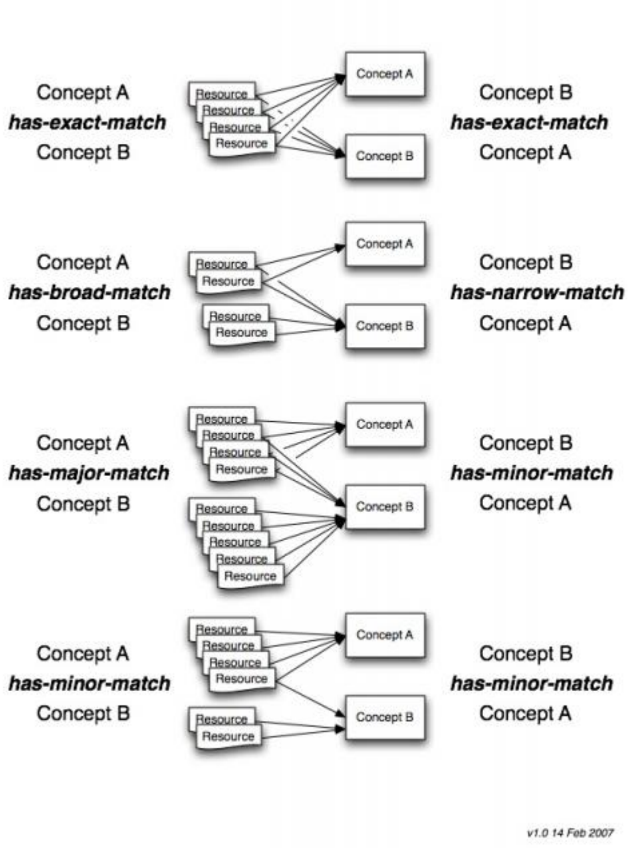
The following basic relationships have been taken from the Simple Knowledge Organization System's (SKOS [210]) Mapping Vocabulary Specification [211]. They offer the ability to distinguish subtle relationships between two terms.

The URI [212] name for each term is shown in brackets, after the label [180] of the term.

```
has-exact-match [exactMatch]
     If two concepts are an exact match, then the set of resources properly indexed against the first concept is identical to
     the set of resources properly indexed against the second.
     Therefore, the two concepts may be interchanged in queries and subject-based indexes. (Is inverse with itself.)


has-broad-match [broadMatch]
     If “concept A has-broad-match concept B,” then the set of resources properly indexed against concept
     A is a subset of the set of resources properly indexed against concept B.
     (Is inverse of has-narrow-match.)

has-narrow-match [narrowMatch]
     If “concept A has-narrow-match concept B,” then the set of resources properly indexed against concept
     A is a superset of the set of resources properly indexed against concept B.
     (Is inverse of has-broad-match.)

has-major-match [majorMatch]
     If “concept A has-major-match concept B,” then the set of resources properly indexed against concept
     A shares more than 50% of its members with the set of resources properly indexed against concept B.
     (No inverse relation can be inferred.)

has-minor-match [minorMatch]
     If “concept A has-minor-match concept B,” then the set of resources properly indexed against concept
     A shares less than 50% but greater than 0 of its members with the set of resources properly indexed against concept B.
     (No inverse relation can be inferred.)
```

Diagram and Discussion of Relationships

The diagram below shows these relations graphically (click for full view).



[213]

Broad and narrow. Be careful using has-broad-match and has-narrow-match, as they may, at first, be counterintuitive. If A *has-broad-match* B, this means (roughly speaking) that B is broader than A, and so A is smaller than B. Since the two relationships, has-broad-match and has-narrow-match, are the inverse of each other, A *has-broad-match* B also implies B *has-narrow-match* A (B is broader than A).

The concepts, has-major-match and has-minor-match, do not have an inverse relationship, as shown in the fourth example in the diagram. It is true that A *has-minor-match* B (because 1 relation mapping to concept A also maps to concept B, but 3 other relations of concept A do not map to column B, so only 1/4=25% of the relations mapping to concept A also map to concept B). But it is also true that B *has-minor-match* A, because only 1 out of 3 of B's relations also map to A.

A similar example could be constructed with has-major-match, where most of the relations in A can be mapped to B, and vice-versa. If both concepts have a has-major-match relation with the other, then the concepts are highly overlapping. If both concepts have a has-minor-match relation, then the concepts overlap a little bit. Finally, if A *has-major-match* B and B *has-minor-match* A, as shown in the third of the four examples, then the A can be considered mostly, but not entirely, a minor subset or part of B.

Unambiguous Web Reference

The generalized terms for these basic relationships, such as “has-broad-match,” are often used within tools or scripts that perform semantic mappings [214]. When using a simple tool like Excel, these relationship terms are sufficient. However, as you begin using your mappings with semantic Web [215] tools and software, it is important to specify relationships in a universally understood syntax.

The Simple Knowledge Organization System (SKOS) Mapping Vocabulary Specification [216] provides a syntax for distributing semantic mapping relationships over the Web. To represent the mapping relationship as a URI (Uniform Resource Identifier), add the prefix "http://www.w3.org/2004/02/skos/mapping#" to the name shown in brackets above. For example, use *http://www.w3.org/2004/02/skos/mapping#narrowMatch* to represent a *has-narrow-match* relationship. URI representation will allow your mapping to be incorporated in the Semantic Web.

Applying the Terms

Deciding whether one thing is the same as another thing in a particular context may not be straightforward. For example, if one data set has measurements of "water temperature" that were all taken at the surface of the ocean, and another data set uses "sea surface temperature," is that an exact match or a broad match? The answer may depend on the context, including the purposes and intended application of the relationships being created. Regardless of context, if the SKOS terms are chosen, a strict application of the definitions should be followed.

In general, developers should use mappings that are accurate in as wide a context as possible, even when considered by someone outside the system, the specific scientific domain, or the broader domain of environmental sciences. The more generally applicable the mappings, the more they can be used by others.

The basic relationship terms above are well defined and understood, and they can be used in a variety of tools and contexts. They provide a basic set of uniquely referenceable relationship terms for use in mapping vocabulary terms.

However, developers may also map data in a limited context with customized relationships defined to convey specific meanings, or they may wish to add other relationships that are specified in a project-specific dictionary [124] of terms or in another standard [157] vocabulary. By using unique Web references for a project's relationship vocabulary, such extensions can be easily created.

Additional Resources

Contact MMI or search in these pages for more information about defining and serving vocabulary terms. The MMI guidance document, "What is 'Same As'?," was created for the MMI vocabulary mapping [217] workshop. It provides more information on defining and using mapping relationships and discusses some of the challenges and pitfalls involved in mapping terms in the real world. Contact MMI and review other guides in the MMI site for more information.

How to cite this Guide

Graybeal, J. 2011. "Semantic Mappings." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/crosswalks/semanticmappings. Accessed June 23, 2014.

Rules for Complex Metadata Mappings

Description of rules in complex mappings

The introduction and definition of rules is an essential step for most cases of creating semantic mapping [214] between standards because of complex relationships [125] that often exist.

To deal with complex mappings (when the mapping from source element [117] to target element is more complex than one-to-one) between standards, we require the introduction of rules.

As an example, consider the case of a source standard having a single element for the address. The target standard may represent the address using multiple elements, such as street address, city, state, zip code, and country. An automated rule could be established to identify certain province or state names, essentially parsing the single element address into its components. Alternatively, a manual rule may also be created, one that specifies that manual intervention is the only method to properly separate the address components.

How to cite this Guide

Kyle, 2011. "Rules for Complex Metadata Mappings." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/crosswalks/rules. Accessed June 23, 2014.

Transformation of Metadata Descriptions

Description of transformation

Transformation [218] is the process of creating a target instance [160] of the metadata [5] description from the source instance. The transformation uses semantic mapping [214] and rules to create the target instance.

It is important to note that the result of the transformation is a metadata description. The created description is sometimes referred to as a crosswalk [219], but this is an inappropriate usage of the word. See the Crosswalk guide for more information about the distinction.

How to cite this Guide

Kyle, 2011. "Transformation of Metadata Descriptions." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/crosswalks/transformation. Accessed June 23, 2014.

Selecting a Standard

Resources to help project managers evaluate and select a standard

Metadata [5] standards are formal specifications [104] of how metadata should be expressed, and following an accepted metadata standard helps ensure that data are appropriately described for later discovery [113] and reuse. But there are many standards, and not all standards are appropriate to a particular project. In choosing a standard from the hundreds available, it is important to evaluate options based on project needs with the goal of creating an interoperable system.

Some criteria for selecting an appropriate standard are presented below, divided into the essential criteria, and some optional additional considerations.

Essential Criteria

Who developed the standard?

The developer will provide an indication of the standard's authority. The most compelling standards tend to be developed by people knowledgeable in a particular discipline or technology. The more collaborative the development process, the greater breadth of understanding went into its initial formation.

Who has implemented and currently uses the standard?

Most standards are implemented by a variety of users. The wider the community of users, the broader the applications for which the standard is likely to be suitable. A broadly used standard is also likely to have additional tools and resources available.

It is not always the case that the most broadly implemented standards are the best for an individual project. From an interoperability [7] standpoint, the most compelling standards are those that are implemented by a variety of users within your scientific area. If multiple organizations have implemented the same standard, then communicating the metadata between organizations in your area can take place with less semantic or syntactic mediation.

Who currently maintains or sponsors the standard?

The sponsor of a standard (that is, the developer or maintainer) affects its authority, relative importance, and acceptance. Many standards are initiated because of a well-documented need within a particular community. These standards tend to be maintained by multiple organizations working collaboratively. For some standards, it is very difficult or impossible to identify who is responsible for maintenance. If there is no clear maintainer, it is likely that the standard will not evolve with the field and may become less and less applicable and useful.

Where is this standard in the development process?

The stage of maturity of a standard is an indication of its development level. Stages of maturity are categorized in the following ways:

- **Missing:** The maturity cannot be determined.
- **Emerging:** The standard is actively being developed—it is in draft, under community review, discussed on mailing lists or forums, mentioned in abstracts, etc.—but has not been formally released.
- **Existing:** The standard is available for public use, has been released or widely adopted, and is sponsored or maintained.
- **Declining:** The standard has less use and is either no longer maintained on a regular basis or is routinely superseded by another emerging or existing standard in the community.

Emerging standards are those undergoing review and first-generation implementation. Projects that implement emerging standards are at the cutting edge. They tend to provide feedback to the developer or maintainer that will result in further development. Well-established, existing standards are ideally accompanied by a dynamic community of users and a variety of resources that can be used in implementation, such as profiles [115], extensions, or vocabularies [130].

Ideally, the metadata manager will choose a standard that is emerging, or existing. Declining standards should be avoided for obvious reasons, and standards for which the maturity level can't be assessed are unlikely to be appropriately documented and maintained. Using an emerging standard may require adjustment over time as it evolves, but if there are no appropriate existing standards, an emerging standard may be the best choice.

Optional Criteria

The optional criteria are much more focused on how implementation of the standard will affect an individual project. These questions represent additional things to consider before implementation.

What is the purpose of this standard?

Each standard is developed for a particular reason. Understanding the reason for the standard's development will provide an indication of how it will benefit, or what it will cost, a particular project. A particular standard may have been created to resolve issues in areas such as metadata format, transmission protocol, limited metadata elements [117], or multiple project objectives. The scenarios below illustrate some of these issues and provide some examples of standards that were created to resolve them.

- Some communities have many metadata formats. While the metadata include the same concepts and terminology, it is nearly

impossible to sift through the varying formats. A standard could prescribe a specific syntax for all metadata files within that specific community (for example, NSDL [220]).
- In some cases, the metadata formats are standardized, but there is not one accepted transmission protocol. One project might submit its metadata via nightly database dumps, while another might use an established Web service [221] to submit metadata in near-real-time. In this case, a technical standard might prescribe a specific methodology for transmitting metadata (for example, Z39.50).
- Some datasets, in their native format, present limited metadata (filename and date, for example). The community of data producers might collaborate on a standard that stipulates required metadata (for example, Marine Geophysical Data Exchange Format - MGD77).
- A project with multiple types of data may have overlapping needs not fulfilled by a particular standard. Implementation of a standard that fulfills multiple objectives needed by a project presents a greater cost-benefit evaluation.

What are the consequences of implementation?

It is often the case that implementation of a standard produces consequences. In some cases, the effects will be positive. In others, the negative effects may outweigh the benefits. Some of these consequences might include the following:

- Compliance with funding agency requirements (positive consequence)
- Interoperability with other projects (positive consequence)
- Need for extensive reorganization and republication (possible negative consequence)

What resources are available for implementation?

While valuable in many ways, standards can be difficult to implement, especially for new users. A standard that is presented with a suite of well-designed tools and resources available for implementation is more compelling than a standard without them.

Types of resources might include:

- Instructional material
- Human support
- Domain-specific profiles
- Software packages to create and/or publish standards-compliant metadata
- Well-developed protocols [132]
- Established controlled vocabularies [122]

Not every data manager will need or want the same set of resources. It is important to know a project's needs and evaluate a standard based on those specific needs.

How to cite this Guide

Neiswender, C., Graybeal, J., Stocks, K. 2013. "Selecting a Standard." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/mdatastandards/standardselect. Accessed June 23, 2014.

Vocabularies: Dictionaries, Ontologies, and More

Introduction to metadata vocabularies, including definitions, basic examples, and links to additional guidance

Every discipline has its own terminology. Consider terms that are used to describe vertical distances. The word “altitude” refers to the distance of something above a reference point like ground level, such as an airplane in flight. If we were examining a set of blueprints for a building we would not use the word “altitude” to describe the level of the rooftop, even though it is also a vertical distance above ground level. Instead, we would use the word “height.” Similarly, if we were in a boat looking down into the water, we would use “depth” rather than “height” to describe the vertical distance.

Also, a single term may be used in multiple communities but with different connotations. For example, an oceanographer may use the term “altitude,” in the operation of a Remotely Operated Vehicle (ROV), to mean the distance above the ocean floor.

In the context of metadata [5], having multiple terms with the same meaning—and terms that can have different meanings in different contexts—can make it harder for people to find and understand data. Using controlled vocabularies [122] within metadata (instead of freely allowing any terminology to be used) can reduce confusion and improve data accessibility.

The Controlled Vocabularies section of the guides describes the importance of controlled vocabularies, the different kinds and their uses, how to implement existing controlled vocabularies, and some considerations for developing new ones.

What is a Controlled Vocabulary?

A vocabulary is a set of terms (words, codes, etc.) that are used in a specific community. In this example, “altitude,” “depth,” and “height” are all part of the vocabularies that scientists and engineers use to talk about vertical distances. It is common for terms to have different connotations in different communities.

A *controlled* vocabulary is a managed set of terms. The management can take different forms, but in controlled vocabularies the allowed terminology is restricted in some way. Within a metadata standard [157] or specification [104], controlled vocabularies are often used to describe the allowed content within a metadata element [117]. This is in contrast to a *free-text metadata element.* As in the example above, in a free-text element, users may choose to use height, altitude, or depth to describe a dataset containing vertical distances. A controlled vocabulary might limit the user's choices—and ensure consistent use of terminology—by specifying that only the term “depth” be used to describe the distance from the ocean’s surface to the seafloor.

For brevity throughout these guides, when we use the term “vocabulary,” we are usually referring to a controlled vocabulary.

Characteristics of a Good Controlled Vocabulary

At a minimum, a controlled vocabulary only needs to manage a set of terms in some way. However, a good controlled vocabulary—one that is easily understood and applied, is likely to be widely adopted, and which improves the clarity of metadata—is one in which the controlled terms are:

- Accepted: the term must adhere to community practices.
- Defined: the terms are precisely characterized; typically, this means the terms have rigorous definitions.
- Managed: experts create, store, and maintain the controlled vocabulary according to agreed-upon procedures. Maintenance involves periodic review, addition of new terms, modification of terms, and occasionally deprecation of terms.

Note that this definition of a controlled vocabulary does not specify a particular scope of usage. Controlled vocabularies could be developed for a local project, for a broader community, or as part of a widely used standard or tool (ISO [128] 19115).

How to cite this Guide

Neiswender, C., Isenor, A., Montgomery, E., Bermudez, L., Miller, S.P. 2011. "Vocabularies: Dictionaries, Ontologies, and More." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs. Accessed June 23, 2014.

What is a Controlled Vocabulary?

Definition of a Controlled Vocabulary

A vocabulary is a set of terms (words, codes, etc.) that are used in a specific community. Vocabularies [130] provide a mechanism for communication- be it written, oral or electronic- because the meaning of the terms are known and agreed upon by the community members. When a vocabulary is formally managed, it becomes a controlled vocabulary. In this case, "managed" means the terms are stored and maintained using agreed-upon procedures. Procedures should exist for adding terms, modifying terms and, more rarely, deprecating terms from a controlled vocabulary.

A controlled vocabulary is a collection of terms that are:

- Accepted: The term must adhere to community practices.
- Defined: The terms are precisely characterized. Typically, this means the terms have rigorous definitions.
- Managed: In general, there will be a body of experts that create and maintain the controlled vocabulary. The controlled vocabulary maintenance will involve periodic review, addition of new terms, modification of terms, and occasionally deprecation of terms.

Notice, this definition does not specify a particular scope of usage. Controlled vocabularies [122] could be developed for a local project (like the Scripps Institution of Oceanography Geological Data Center), a broader community (e.g. OOSTethys), or as a part of a widely used standard [157] or tool (ISO [128] 19115).

Controlled Vocabulary Categories and Types

To many people, the English language is a well-known vocabulary. We have many ways of representing the terms in the English language. For example, if we want to figure out what a specific word means we might consult a glossary [13]; if we want to know the origin of a term we might consult a dictionary [124]; and if we want to know how the term relates to other terms we might consult a thesaurus [105]. We also need to recognize that the meaning of terms may change through time. Generations use terms in different ways (cool in one generation means a low temperature, while cool in another is a positive adjective).

To enable formal management, a controlled vocabulary can be organized in several ways. There are three broad categories of controlled vocabularies: flat, multi-level and relational.

- Flat controlled vocabularies provide a set of used terms. Some flat controlled vocabularies will provide additional information about each term.
- Multi-level controlled vocabularies build upon a flat controlled vocabulary by assigning each term to a category.
- Relational controlled vocabularies provide a set of terms, and capture how they are associated with each other.

Within these three categories, there are additional controlled vocabulary types. The table below summarizes these categories and types. The table categorizes necessary conditions only. Some controlled vocabularies will appear as "hybrids" of one or more categories of controlled vocabularies. Please see the Types of Controlled Vocabularies [42] guide for a more extensive explanation, or this article on Knowledge Organization Systems [222].

Broad Category	Controlled Vocabulary Types	Description
Flat Controlled Vocabulary	Authority File [223]	List of terms
	Glossary	List of terms and definitions within a specific domain
	Dictionary	List of terms, definitions, and additional information
	Gazetteer [106]	List of place names
	Code List [196]	List of codes (e.g. abbreviations) and definitions
Multi-Level Controlled Vocabulary	Taxonomy [224]	Terms classified into categories
	Subject Heading [186]	Terms classified into categories, which may be broad classes [225]
Relational Controlled Vocabulary	Thesaurus	Set of terms and relationships [125] among individual values [119]
	Semantic Network [226]	Set of terms/concepts and directed relationships
	Ontology [107]	Set of terms and relationships among terms, enhanced by additional information provided by rules and axioms.

The Purpose of a Controlled Vocabulary

Controlled vocabularies can serve several different purposes. For example, a controlled vocabulary might help users find data (also known as a "discovery vocabulary [113]"), or assist in the interpretation of data (also known as a "usage vocabulary [123]"). The controlled vocabulary might provide human-understandable meaning (also known as a "semantic vocabulary") or machine-readable [112] format information (also known as a "syntactic vocabulary"). Controlled vocabularies provide these abilities by:

- establishing the permissible terms to be used;
- maintaining the proper and agreed-upon spelling of the terms;
- clarifying terms for those who are new to the community; and
- eliminating the use of arbitrary terms that can cause inconsistencies and confusion.

How to cite this Guide

Neiswender, C. 2009. "What is a Controlled Vocabulary?." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/vocdef. Accessed June 23, 2014.

The Importance of Controlled Vocabularies

Description of how a controlled vocabulary can enhance a system

Controlled Vocabularies [122] are important to researchers for many reasons:

- Consistency
- Accuracy
- Automation
- Simplification of input
- Interoperability [7]
- Enhancement of searches and discovery [113]
- Completeness
- Long- and short-term management
- Efficient use of time

In many cases, controlled vocabulary terms [209] completely define the allowable content for a particular metadata [5] 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 [117], 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.

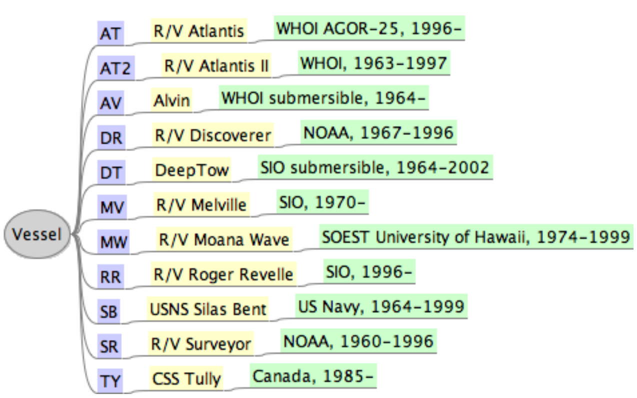
The different types of controlled vocabularies provide different levels of interoperability. Often when we move from one project to another, we need to identify the metadata descriptions that use one controlled vocabulary and relate these descriptions to another system. We might need to understand more about the terms in the initial controlled vocabulary: what it represents (glossary [13]), how it came to be (dictionary [124]), and what terms are similar (thesaurus [105], semantic network [226], 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 [186], taxonomies [224], and ontologies [107]) 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.

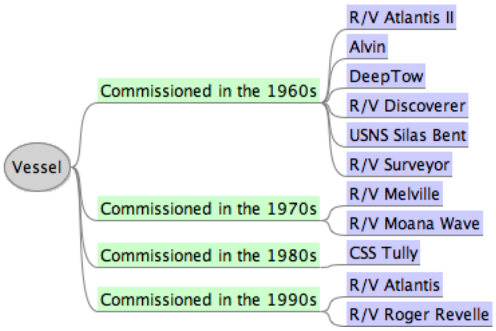
Figure 1



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)

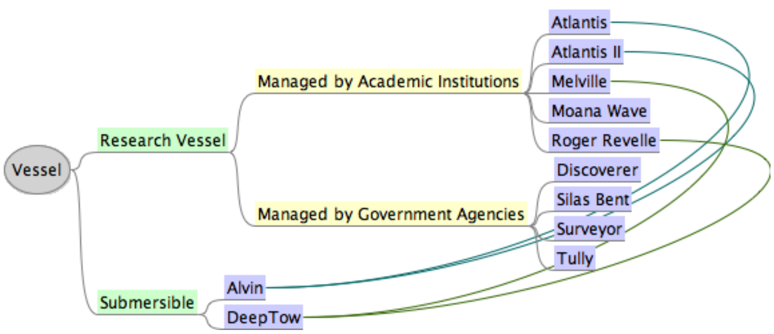
Figure 2



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 [225] (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 [119] 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.

How to cite this Guide

Bermudez, L., Montgomery, E., Miller, S.P., Neiswender, C., Isenor, A. 2011. "The Importance of Controlled Vocabularies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/vocimportance. Accessed June 23, 2014.

Classification of Controlled Vocabularies

Metadata vocabularies described by their form (category) and function (type)

In understanding English, if we want to figure out what a word means, we might consult a dictionary [124] or a glossary [13]. Or we may use an etymology dictionary to track the history [108]

The record of how a particular value or record came to be. Provenance can include things like when, by whom, and how the item was created or modified.

of a word. If we want to know how a term relates to other terms we might consult a thesaurus [105].

Like the vocabulary sources for the English language, controlled vocabularies [122] for describing metadata [5] can be classified by their purpose, their form, or their functionalities.

Classification by Purpose

Vocabularies may be defined by their ability to accomplish specific goals:

- Discovery vocabulary [113]: helps users find data
- Usage vocabulary [123]: assists in the interpretation of data
- Semantic vocabulary: provides human-understandable meaning
- Syntactic vocabulary: translates information into machine-readable [112] format

Controlled vocabularies provide these abilities by

- establishing the permissible terms to be used;
- maintaining the proper and agreed-upon spelling of the terms;
- clarifying terms for those who are new to the community; and
- eliminating the use of arbitrary terms that can cause inconsistencies and confusion.

Classification by Form

To enable formal management, a controlled vocabulary can be organized structurally such that that it fits into one of these broad categories:

- Flat: provides a set of required terms that may be used. Some flat controlled vocabularies will provide additional information about each term.
- Multilevel: builds upon a flat controlled vocabulary by assigning each term to a category.
- Relational: provides a set of terms and captures how they are associated with each other.

Classification by Functionality

Within the three broad categories that classify controlled vocabularies by form, there are sub [227]-groupings that we will call "types." The table below summarizes the relationships [125] between the broad, form-based categories and their respective function-based types. The table defines the types and categories according to their minimum required characteristics.

Broad, Form-based Category	Functionality-based Type	Description
Flat Controlled Vocabulary	Authority File [223]	List of terms
	Glossary	List of terms and definitions within a specific domain
	Dictionary	List of terms, definitions, and additional information
	Gazetteer [106]	List of place names
	Code List [196]	List of codes (e.g., abbreviations) and definitions
Multilevel Controlled	Taxonomy [224]	Terms classified into categories

Vocabulary	Subject Heading [186]	Terms classified into categories, which may be broad classes [225]
Relational Controlled Vocabulary	Thesaurus	Set of terms and relationships among individual values [119]
	Semantic Network [226]	Set of terms/concepts and directed relationships
	Ontology [107]	Set of terms and relationships among terms, enhanced by additional information provided by rules and axioms.

Hybrid Classifications and the Real World

Not all controlled vocabularies fit neatly into one type; some may appear as hybrids or crossovers. Vocabularies rarely exist in a vacuum and evolve over time, causing the distinctions between the classifications to be muddied, either intentionally or unintentionally. In addition, vocabularies can fit multiple classifications.

Consequently, one controlled vocabulary might fit the definition of more than one type. For example, an ontology might also have many of the characteristics of a dictionary. Because of this ambiguity, the different types may be referred to generically as "vocabularies" or "controlled vocabularies," especially if they have hybrid characteristics.

Comparing and Understanding Classifications of Controlled Vocabularies

The guides in this section contain several articles to help you understand the distinction between classifications of controlled vocabularies and to examine some types side by side. Also, see the article Knowledge Organization Systems [222] for more information.

How to cite this Guide

Neiswender, C., Isenor, A., Montgomery, E., Bermudez, L., Miller, S.P. 2011. "Classification of Controlled Vocabularies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes. Accessed June 23, 2014.

Usage vs Discovery Vocabularies

Introduction to usage and discovery vocabularies, definitions and examples

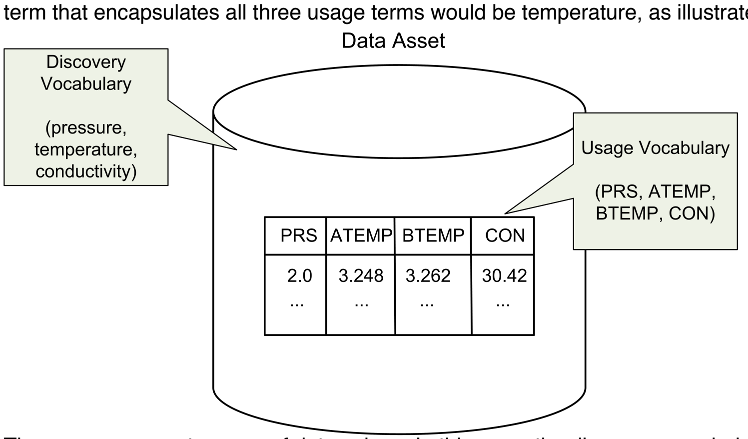
In the guide Vocabularies [130]: Dictionaries [124], Ontologies [107], and More, we used the term “altitude” to describe part of the spatial position of something. We may complete the spatial description by including the terms “latitude” and “longitude.” “Latitude” typically refers to a value that describes north-south placement (or y-coordinate) of something on the earth (more generally a rotational ellipsoid). Used with the term “longitude” (to describe the east-west placement (or x-coordinate), we can fully specify the position of something on the earth.

Consider a data asset that contains altitude, latitude, and longitude values [119]. The asset may be a database table, a spreadsheet, or a text file. The asset could use plain English names for the columns of numbers, that is, it could use altitude, latitude and longitude. Alternatively, the names could be cryptic codes or abbreviations such as ALT, LAT, and LONG. The names used within the asset represent what we refer to as a usage vocabulary [123]. A usage vocabulary is important when clients—or software applications—want to effectively access the data.

However, when discovering the content of an asset, the usage vocabulary may or may not be useful depending on how cryptically the data columns are named.

Therefore, to facilitate discovery, we use a discovery vocabulary [113], which uses terminology to identify the data that are common to the subject community. Discovery vocabulary terms [209] can take a variety of forms:

- They may be identical to terms in the usage vocabulary. This is the situation when the data asset uses common language terminology to identify the data, for example, data values identified as temperature or salinity.
- They may represent groups of terms in the usage vocabulary. This is a common situation for legacy assets, where cryptic codes have been used to identify similar data from multiple sources. For example, consider a legacy data asset that contains temperature values from sensors A, B, and C. These data are identified within the asset as ATEMP, BTEMP, CTEMP. The discovery vocabulary term that encapsulates all three usage terms would be temperature, as illustrated in the image below.



- They may represent groups of data values. In this case, the discovery vocabulary terms identify particular subgroups of the data, rather than all of the data. For example, if the data asset contains geology data, then certain geological time periods (e.g., Mesozoic Era) may be identified in the discovery vocabulary. In physical oceanography, a discovery term may identify a particular water mass (e.g., Labrador Sea Water) that has particular characteristics (e.g., physical or chemical).

Discovery vocabularies aid a person in finding the data asset, while the usage vocabulary aids in use of the asset. Both vocabularies can pertain to data-related topics such as parameters, platforms, sensors, geographic areas, etc., and both usage and discovery vocabularies are specialized forms of controlled vocabularies [122].

How to cite this Guide

Isenor, A. 2011. "Usage vs Discovery Vocabularies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/cvusagevsdisc. Accessed June 23, 2014.

Metadata and Vocabularies

Metadata vs. Vocabularies

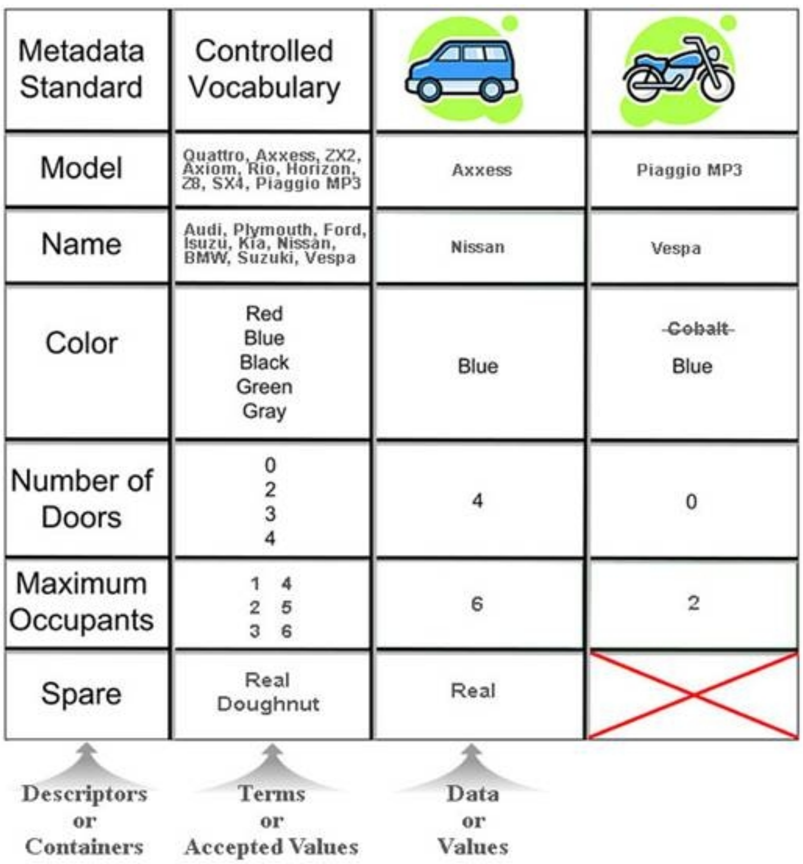
Metadata [228] is used to describe the aspects of something. In the MMI [228] community, the item being described could be almost anything related to the marine community, such as a data set or a marine service.

A metadata standard [229], also known as a content standard [230], is used to define the containers for the metadata. It is a list of possibly important descriptors for the item. Since the list contains many possible containers (descriptors), any specific container may or may not be of use for a particular item being described. A controlled vocabulary [230] is often used as the content for specific metadata containers. The controlled vocabulary, which is a managed list of acceptable terms, will determine all the possible values [119] that may be used in the container. If someone fills the container with text that is not included in the accepted values, it will not be permitted to stay in the container.

As an example, the diagram below shows the relationship [125] between the metadata descriptors in the metadata standard, the terms in the controlled vocabulary, and the data.

Suppose we are going to purchase a vehicle. In this example, the vehicle represents the item we are going to describe. A hypothetical metadata standard to assist us in vehicle selection might consist of an extensive list of important containers that describe any possible vehicle. These containers could include model name, color, number of doors, number of passengers, and type of spare tire.

Note that the descriptors—the containers—in the list do not apply to every possible vehicle. For example, a motorcycle has no spare tire, and thus the descriptor for spare does not apply.



Metadata Standard	Controlled Vocabulary		
Model	Quattro, Axxess, ZX2, Axiom, Rio, Horizon, Z8, SX4, Piaggio MP3	Axxess	Piaggio MP3
Name	Audi, Plymouth, Ford, Isuzu, Kia, Nissan, BMW, Suzuki, Vespa	Nissan	Vespa
Color	Red Blue Black Green Gray	Blue	~~Cobalt~~ Blue
Number of Doors	0 2 3 4	4	0
Maximum Occupants	1 4 2 5 3 6	6	2
Spare	Real Doughnut	Real	

Consider the Number of Doors container. In this case, the controlled vocabulary may be represented by the numbers 0, 2, 3, and 4. This controlled vocabulary allows for all passenger vehicles produced by all manufacturers. The zero case accounts for motorcycles, the two-door case for two-door cars, and the three- and four-door case for the older and newer style vans. No other value [231] for the number of doors is possible in passenger vehicles.

When using a controlled vocabulary, the values chosen for the data must conform to the terms in the controlled vocabulary. In this example, the color of the Vespa might be called cobalt by the manufacturer, but because we are using a controlled vocabulary, we must choose an accepted term. Cobalt is not an accepted value, so we need to choose one from the accepted list, in this case, blue.

In this example, the metadata standard represents the structure and metadata descriptors, while the controlled vocabulary is used to define the allowable content (e.g., the 0, 2, 3, 4 list) for the descriptors.

How to cite this Guide

Isenor, A. 2011. "Metadata and Vocabularies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/metadataandcv. Accessed June 23, 2014.

Semantic vs. Syntactic Vocabularies

Semantic vs. Syntactic Vocabularies

Semantics is the meaning of words. Semantic vocabularies [130] provide meaning to the terms used in metadata [5] in a way that is understandable to a human being. For example, the semantic vocabulary description for *altitude* might be, "the vertical position of a flying object."

Syntax in metadata refers to the format instructions for storage of values [119] in computers. Syntax might include information on values such as float, real, ASCII [232] or binary. For example, an entry in a syntactic vocabulary might be, "altitude data values are measured in feet, F8.6 (here, the F8.6 indicates the form of the numeric value, 8 characters in total with the possibility of 6 after the decimal)."

As noted previously, usage vocabularies [123] provide information on the terminology for using the data values. Thus, a usage vocabulary would include both semantic metadata (e.g., the term *latitude* and the definition of what this term means) and syntactic metadata (e.g., the data value for latitude is F8.6).

However, a discovery vocabulary [113] typically does not contain syntactic metadata. This is because the discovery vocabulary describes collections of usage terms or data values that already are known to the community.

For example, *North Atlantic Ocean* represents a certain grouping of latitude and longitude values and represents a specific meaning to the oceanographic community. The term *North Atlantic Ocean* has inherent syntax (e.g., capitalization and allowed spaces) but does not have syntax associated with the content because the term doesn’t explicitly contain values but rather is a generalized description of a collection of values.

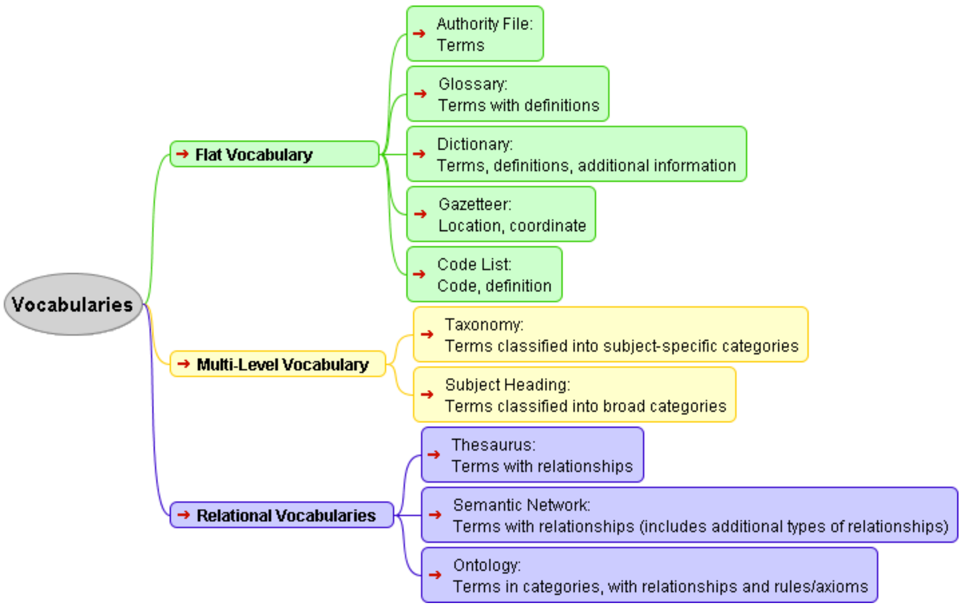
How to cite this Guide

Isenor, A. 2011. "Semantic vs. Syntactic Vocabularies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/cvsemvssyn. Accessed June 23, 2014.

Categories of Controlled Vocabularies

Overview of the categories of vocabularies, including flat, multi-level and relational vocabularies

As defined in Classification of Controlled Vocabularies [122], vocabularies can be defined by their structure and form. There are three broad categories of controlled vocabularies: flat, multilevel, and relational (also called a relationship list). Within these three categories, there are a variety of types of controlled vocabularies. A *type* is a simplified name for vocabularies further classified by function. (Click on image to enlarge.)



Flat Vocabularies

Authority File [43] * Glossary [14] * Dictionary [44] * Gazetteer [45] * Code List [46]

All flat vocabularies contain a label [180] and a value [119]. Some flat vocabularies build upon this foundation by adding a definition or additional information about each value. No relationships [125] are established, no hierarchies are set up, and no complicated matrices are necessary. Flat vocabularies are sets of two to four pieces of information: a label, a value, and possibly a definition and additional information.

Multilevel Vocabularies

Taxonomy [47] * Subject Heading [48]

A multilevel vocabulary is essentially a way to group terms into classes [225] with hierarchy. A classification tells more about the terms by placing them into well thought-out subcategories.

Think of a classification as a tree with a trunk, limbs, branches, and leaves. If you look at an individual leaf on the tree, you can backtrack to the branch, to the limb, and eventually to the trunk.

In a multilevel vocabulary, you can examine in which subcategory a term belongs, and you can examine the relationships between subcategories as well. In some multilevel vocabularies, the only connection between the subcategories is a broader than/narrower than comparison (taxonomy). In others, you can compare similar categories across broader categories (subject heading).

Relational Vocabularies

Thesaurus [49] * Semantic Network [50] * Ontology [51]

Relational Vocabularies, also called relationship lists, contain a mechanism to connect terms. The relations are described by various standards and protocols [132], such as for thesauri [105] in the ANSI [233]/NISO [109] Z39.19 - 2005 standard, including broader than/narrower than, used for, and related.

The principles of a relationship list can be illustrated by seashells. You might find one on the San Diego coast that looks exactly like one in Monterey Bay, except the shell in Monterey Bay has been degraded by extreme waves and the San Diego shell has not. These two are probably related (similar to *broader than*). Or, perhaps you see a shell on the Oregon coast, and it looks similar to a shell in Woods Hole, Massachusetts. However, the two aren't exactly the same, because they're different colors. There is still a relationship to be defined (similar to *related*). Perhaps you find a message in a bottle in Maryland. The bottle came from Europe, and there's a seashell inside. In this case, you would want to relate the shell to other European shells (similar to *used for*).

How to cite this Guide

Neiswender, C., Isenor, A. 2011. "Categories of Controlled Vocabularies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/voccat. Accessed June 23, 2014.

Authority File (Flat Vocabulary)

Definition, description of authority files

Definition, description of authority files

Flat Vocabularies [234]					Multi -Level Vocabularies		Relational Vocabularies [236]		
Authority File	Glossary [238]	Dictionary [239]	Gazetteer [240]	Code List [241]	Taxonomy [242]	Subject Heading [243]	Thesaurus [244]	Semantic Network [245]	Ontology [246]

Definition of Authority File

A type of flat controlled vocabulary [122] consisting of a list of labels and values that establish the acceptable values that can be inserted into a particular parameter.

Discussion

No explanation or augmenting information is given about the acceptable values. To implement an authority file, project managers must have a clear understanding of both the metadata parameter and the domain in which the authority file is applicable.

Example - Dublin Core Metadata Initiative (DCMI) [247]

```
<xs:simpleType name="DCMIType">

    <xs:union>

        <xs:simpleType>

            <xs:restriction base="xs:Name">

                <xs:enumeration value="Collection"/>
                <xs:enumeration value="Dataset"/>
                <xs:enumeration value="Event"/>
                <xs:enumeration value="Image"/>
                <xs:enumeration value="InteractiveResource"/>
                <xs:enumeration value="Service"/>
                <xs:enumeration value="Software"/>
                <xs:enumeration value="Sound"/>
                <xs:enumeration value="Text"/>
                <xs:enumeration value="PhysicalObject"/>

            </xs:restriction>

        </xs:simpleType>

    </xs:union>

</xs:simpleType>
```

Notice, to implement this appropriately, you would need to know that DCMI stands for Dublin Core Metadata Initiative, and Type is one of the 15 required metadata parameters. Dublin Core can be widely implemented, as it contains very broad parameters and values. This particular Authority File is presented in XML [131] format. It was distributed in this format via the web by the managing body.

How to cite this Guide

Neiswender, C. 2011. "Authority File (Flat Vocabulary)." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/voccat/authorityfile. Accessed June 23, 2014.

Glossary (Flat Vocabulary)

Definition, description of a glossary

Definition, description of a glossary

Flat Vocabularies [234]					Multi -Level Vocabularies		Relational Vocabularies [236]		
Authority File [248]	Glossary	Dictionary [239]	Gazetteer [240]	Code List [241]	Taxonomy [242]	Subject Heading [243]	Thesaurus [244]	Semantic Network [245]	Ontology [246]

Definition of Glossary

A type of flat controlled vocabulary [122] containing a list of terms in a particular domain of knowledge with the definitions for those terms.

Discussion

A glossary builds upon an authority file. For each value [119], a glossary provides users with a description that articulates exactly what each value means. The vocabulary is a triplet: label [180], term (accepted value), and description. This adds functionality for system users, compared to the authority file, because abstract terminology becomes more concrete.

Example - SIO/GDC Vocabulary Terms [249]

- "DCMI [190]_Format","Dataset","General Digital Data set"
- "DCMI_Format","Image","Photographic resource, such as a jpg or tiff"
- "DCMI_Format","InteractiveResource","Enhanced data set, which allows for user input"
- "DCMI_Format","PhysicalObject","Physical sample, paper document, or digital media"
- "DCMI_Format","Software","Software code"
- "DCMI_Format","Sound","Acoustic Recording, such as an audio tape or file"
- "DCMI_Format","Text","General textual resource, such as Word, ASCII [232], or Acrobat "

In this example, the vocabulary builds upon the DCMI authority file. In this glossary, the label (also called a parameter [117]) is DCMI_Format, which correlates with the Dublin Core Type. The SIO/GDC Vocabulary Terms use a subset of the complete DCMI list, and formats them as a comma-separated-value (CSV [135]) text file. This helps to eliminate incorrect implementation, and facilitates understanding for users.

How to cite this Guide

Neiswender, C. 2011. "Glossary (Flat Vocabulary)." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/voccat/glossary. Accessed June 23, 2014.

Dictionary (Flat Vocabulary)

Definition, description and example of a dictionary

Definition, description, and example of a dictionary

Flat Vocabularies [234]					Multi -Level Vocabularies		Relational Vocabularies [236]		
Authority File [248]	Glossary [238]	Dictionary	Gazetteer [240]	Code List [241]	Taxonomy [242]	Subject Heading [243]	Thesaurus [244]	Semantic Network [245]	Ontology [246]

Definition of Dictionary

In the context of metadata [5], a dictionary is a type of controlled, flat vocabulary that provides a list of metadata terms, definitions, and additional information within a specific domain.

Discussion

In practice, a dictionary builds upon a glossary by providing additional information about each term, or value [119]. This additional information might include how the value came to be selected, the etymology of the value, or additional, site-specific context for each value. A dictionary provides more than a definition, but the nature of the additional information is subject to each project's interpretations.

Example - CDI Sea Search (SeaDataNet) Vocabularies [250]

- "CODE","SHORT_TITLE","FULL_TITLE","COMMENTS","CREATED","MODIFIED"
- "USPC","Unspec.","Not specified","Use for parameters where the units are deliberately undefined such as arbitrary units","2/2/05 16:14",""
- "UKWN","Unknown","Not known","Use for parameters that should have a unit, but it is not known. Usage of this code is STRONGLY DISCOURAGED","2/2/05 16:16",""
- "UPPM","ppm","Parts per million","Usage not recommended for parts per million by weight. Use milligrams/kilogram instead.","16/9/1994 00:00:00","18/10/1995 13:36:11"
- "UAQU","A^2/Q","Angstrom squared per quanta","Scaled unit by Avagadro's number and powers of ten to m**2/mol","2/2/05 16:49",""
- "USVD","Sv","Sverdrup","One Sverdrup is a million cubic metres per second or a million cumecs","17/2/2005 12:44:16","14/10/2005 11:47:50"
- "UNPI","ng","Nanograms","Changed from nanograms per individual when per individual semantics transferred to the parameter description","10/2/00 0:00","3/8/05 8:52"
- "UAAA","deg","Degrees","","1/1/87 0:00","4/11/05 12:47"
- "UABB","deg T","Degrees True","","1/1/87 0:00","4/11/05 12:47"

This dictionary provides a code, a short title, a full title, comments, creation date, and modification date, in csv [135] format. It clearly tells us about the management of this vocabulary, which is very useful for selecting a vocabulary [52].

How to cite this Guide

Neiswender, C. 2011. "Dictionary (Flat Vocabulary)." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/voccat/dictionary. Accessed June 23, 2014.

Gazetteer (Flat Vocabulary)

Definition, description and example of a gazetteer

Definition, description, and example of a gazetteer

Flat Vocabularies [234]					Multi -Level Vocabularies		Relational Vocabularies [236]		
Authority File [248]	Glossary [238]	Dictionary [239]	Gazetteer	Code List [241]	Taxonomy [242]	Subject Heading [243]	Thesaurus [244]	Semantic Network [245]	Ontology [246]

In the context of metadata [5], a gazetteer is a geographic term list, which is a specific type of flat controlled vocabulary [122].

Discussion

The main point of a gazetteer is to identify locations within a standardized coordinate system. For systems to be able to communicate effectively, locations must be clearly articulated.

The additional information provided by a gazetteer includes the necessary points of reference to find that particular location. These points of reference take the form of coordinates (lat/lon, x/y, etc.); the coordinates themselves come from a taxonomy that provides coordinates for a list of pre-defined locations. (Taxonomies [224] are discussed in the Multi-Level Vocabularies [251] section.)

In a global environment this is very important. If a project divides the globe into a set of very specific regions, the gazetteer provides reference points and, therefore, meaning to the names for these regions. By necessity, a gazetteer will encompass a GIS [129] (geographic information system). It can take the form of an XML [131] file, an ArcGIS layer, a shapefile, or a location information file designed for a home-grown interface.

Example - National Geospatial Intelligence Agency Country Files [253]

Click on image to enlarge.

RC	UFI	UNI	LAT	LONG	DMS_LAT	DMS_LONG	MGRS	JOG	FC	DSG	PC	CC1	ADM1	ADM2	POP	ELEV	CC2	NT	LC	SHORT_FORM	GENERIC	SORT_NAME	FULL_NAME	FULL_NAME_ND	MODIFY_DATE
1	-1307690	-1891634	12.533333	-70	123200	-700000	19PCP9135585735	ND19-14	H	STMI		AA	00					N				ROOIAFO	Rooi Afña	Rooi Afo	1993-12-21
1	-1307691	-1891635	12.45	-69.933333	122700	-695600	19PCP9856676493	ND19-14	T	HLL		AA	00					V				CERUALEJANDRO	Ceru Alejandro	Ceru Alejandro	1993-12-21
1	-1307691	-1891636	12.45	-69.933333	122700	-695600	19PCP9856676493	ND19-14	T	HLL		AA	00					N				SERIALEJANDRO	Ser,Áõi Alejandro	Serí Alejandro	1993-12-21
1	-1307692	-1891637	12.566667	-70.016667	123400	-700100	19PCP8955889429	ND19-14	T	HLL		AA	00					V				ALTOVISTA	Alto Vista	Alto Vista	1995-03-23
1	-1307692	60432	12.566667	-70.016667	123400	-700100	19PCP8955889429	ND19-14	T	HLL		AA	00					N			Ceru	CERUALTOVISTA	Ceru Alto Vista	Ceru Alto Vista	1995-03-23

Click on image to enlarge.

This gazetteer (represented in a table for web display purposes) is available online. Each area is completely described in one row of the document. Documentation provided by the managing body is extremely important. Without it, you might be able to guess that LAT is latitude in decimal degrees, but you might not understand that DMS_LAT is latitude in degrees, minutes, and seconds. In addition, codes used throughout this gazetteer indicate things like a region or a feature classification. Neither the gazetteer nor the code list would make sense alone.

How to cite this Guide

Neiswender, C. 2011. "Gazetteer (Flat Vocabulary)." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/voccat/gazetteer. Accessed June 23, 2014.

Code List (Flat Vocabulary)

Definition, description and example of a code list

Definition, description and example of a code list

Flat Vocabularies [234]					Multi -Level Vocabularies		Relational Vocabularies [236]		
Authority File [248]	Glossary [238]	Dictionary [239]	Gazetteer [240]	Code List	Taxonomy [242]	Subject Heading [243]	Thesaurus [244]	Semantic Network [245]	Ontology [246]

Definition of Code List

A code list is a type of flat controlled vocabulary [122] consisting of a set of codes and meanings used in a specific project.

Discussion

Like the gazetteer, a code list is a very specific type of vocabulary that allows users to interpret metadata [5] whose labels are abbreviated or represented as numerals or acronyms. The codes could be computer-readable, human-readable, or a hybrid of each.

Example - GCMD Sensors Vocabulary [254] (Instrument Keywords)

- AIRS > Atmospheric Infrared Sounder
- AIS > Arizona Imager Spectrograph
- AMI > Active Microwave Instrument
- ATM > Airborne Topographic Mapper
- BIONESS > Bedford Institute of Oceanography Net Environmental Sampling System
- CAPS > Cloud, Aerosol, Precipitation Spectrometer
- CERES > Clouds and Earth's Radiant Energy System
- CPA > Charged Particle Analyzer
- CPC > Cloud Particle Counter
- CRS > Cloud Radar System
- DIAL > Differential Absorption Lidar
- EPIC > Energetic Particle and Ion Composition (Geotail)
- ERB > Earth Radiation Budget
- GLAS > Geoscience Laser Altimeter System
- GOLF > Global Oscillations at Low Frequencies
- GSD > Germanium Semiconductor Detector
- KBR > K-Band Ranging system
- LIP > Lightning Instrument Package
- LIS > Lightning Imaging Sensor
- MACAWS > Multicenter Airborne Coherent Atmospheric Profiling Wind Sensor
- MBLA > Multi-Beam Laser Altimeter
- MWA > Multiple Water Analyzer
- NMLR > Nebraska Multiband Leaf Radiometer
- OPC > Optical Plankton Counter
- ORG > Optical Rain Gauge
- PALS > Passive and Active L- and S-Band System
- PARABOLA > Portable Apparatus Rapid Acquisition Bidirectional Observ Land and Atmos
- ROWS > Radar Ocean Wave Spectrometer
- SAR > Synthetic Aperture Radar
- SLAR > Side-Looking Airborne Radar
- SOPA > Synchronous Orbit Particle Analyzer
- SOUP > Solar Optical Universal Polarimeter
- TEAMS > Time of Flight Energy Angle Mass Spectrometer
- USO > Ultra-Stable Oscillator
- WFC > Wide Field [117] Camera
- WS > Wind Scatterometer

This list of codes represents human-readable codes and what they mean. This list is formatted in a very human-friendly way: code > definition.

Example - CDI Sea Search (SeaDataNet) Vocabularies [250] (FTP Site - seasearch_category.csv)

- "CODE","TITLE"
- "B005","Bacteria and viruses"
- "B007","Biota composition"

- "B015","Birds, mammals and reptiles"
- "B020","Fish"
- "B025","Microzooplankton"
- "B027","Other biological measurements"
- "B030","Phytoplankton"
- "B035","Pigments"
- "B040","Zoobenthos"
- "B045","Zooplankton"
- "C003","Amino acids"
- "C005","Carbon, nitrogen and phosphorus"
- "C010","Carbonate system"
- "C015","Dissolved gases"
- "C017","Fatty acids"
- "C020","Halocarbons (including freons)"
- "C025","Hydrocarbons"
- "C030","Isotopes"
- "C035","Metal concentrations"
- "C040","Nutrients"
- "C045","Other inorganic chemical measurements"
- "C050","Other organic chemical measurements"
- "C055","PCBs and organic micropollutants"
- "D005","Acoustics"
- "D015","Optical properties"
- "D025","Water column temperature and salinity"
- "D020","Other physical oceanographic measurements"
- "D030","Currents"
- "D032","Sea level"
- "D034","Waves"
- "G005","Gravity, magnetics and bathymetry"
- "G010","Sediment properties"
- "G012","Sonar and seismics"
- "G015","Suspended particulate matter"
- "G020","Sedimentary rock geochemistry and geochronology"
- "G025","Igneous and metamorphic rock geochemistry and geochronology"
- "M005","Atmospheric chemistry"
- "M010","Meteorology"
- "M015","Cryosphere"
- "O005","Fluxes"
- "O010","Rate measurements (including production, excretion and grazing)"
- "T001","Terrestrial"
- "Z005","Administration and dimensions"

In this example, the codes are largely computer-readable. Originally formatted as comma-separated values [119] (seasearch_categories.csv) in a text file, this code list provides a human-understandable interpretation of computer-readable codes.

How to cite this Guide

Neiswender, C. 2011. "Code List (Flat Vocabulary)." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/voccat/codelist. Accessed June 23, 2014.

Taxonomy (Multi-Level Vocabulary)

Definition, description and example of a taxonomy

Definition, description and example of a taxonomy

Flat Vocabularies [234]					Multi -Level Vocabularies		Relational Vocabularies [236]		
Authority File [255]	Glossary [256]	Dictionary [257]	Gazetteer [258]	Code List [259]	Taxonomy	Subject Heading [260]	Thesaurus [261]	Semantic Network [262]	Ontology [263]

Definition of Taxonomy

A multilevel controlled vocabulary [122] in which metadata values [119] are grouped according to subject-specific classes [225], usually hierarchical.

Discussion

Taxonomies [224] are an organizational structure in which metadata values are grouped according to subject-specific classes. Each class has a specific description, which is the set of characteristics that each member of a class exhibits. Taxonomies begin with the broadest of classes, and continue to narrow until the final class is reached.

The classes that are included in a taxonomy will depend largely upon the discipline of the metadata values. So, while a geological taxonomy might include classes like metamorphic, igneous and sedimentary, a chemical taxonomy might include classes like solid, liquid and gas.

These classes provide characteristics for each of the members of a class. The taxonomy is used to describe a particular thing, and usually culminates in a unique identifier for each member.

Note: While taxonomies tend to include subject-specific, narrow classes, another type of classification, subject headings [186], tends to include broader classes.

Example

Perhaps the most well-known taxonomy is the Linnaean taxonomy, which uniquely classifies living things. In its simplest form, the Linnaean taxonomy has seven main levels. As an example, consider the Linnaean classification for humans:

- **Kingdom:** Animalia

  Characteristic: eukaryotic cells with a cell membrane but no cell wall, multi-cellular, heterotrophic

- **Phylum:** Chordata

  Characteristic: notochord, dorsal nerve cord, and pharyngeal gill slits

- **Class:** Mammalia

  Characteristic: endothermic, hair and mammary glands (which are used to nourish the young)

- **Order:** Primates

  Characteristic: collar bone, eyes facing forward, grasping hands with fingers, two types of teeth

- **Family:** Hominidae

  Characteristic: upright posture, large brain, stereoscopic vision, flat face, hands and feet with different functions

- **Genus:** Homo

  Characteristic: s-curved spine

- **Species:** sapiens

  Characteristic: high forehead, well-developed chin, skull bones thin

Notice, as human beings, we are given a unique identifier (Homo sapiens)*, and we exhibit all the characteristics listed (in other words, since our classification is at the bottom of a nested list, we can inherit all the characteristics of the "super-classes"). To completely classify humans in this taxonomy, we need to use the term* Homo sapiens*, but you could also call human beings primates. This would not be the narrowest classification, but it is an accurate classification.*

How to cite this Guide

Neiswender, C. 2011. "Taxonomy (Multi-Level Vocabulary)." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/voccat/taxonomy. Accessed June 23, 2014.

Subject Heading (Multi-Level Vocabulary)

Definition, description and example of a subject heading

Definition, description and example of a subject heading

Flat Vocabularies [234]					Multi -Level Vocabularies		Relational Vocabularies [236]		
Authority File [248]	Glossary [238]	Dictionary [239]	Gazetteer [240]	Code List [241]	Taxonomy [242]	Subject Heading	Thesaurus [244]	Semantic Network [245]	Ontology [246]

Definition of a Subject Heading

A type of multilevel controlled vocabulary [122] in which metadata [5] values [119] are classified into categories that may be broad classes [225].

Discussion

Subject headings [186] provide an organizational structure in which metadata values are grouped according to broad classes. This type of vocabulary provides a big picture scheme, where values can be viewed in the context of established, overarching subjects.

The classes in a subject heading are used to provide a contextual basis for each member; that is, the subject heading shows how a particular value fits into an established context. A particular value can be used in multiple contexts, so a value can be found in multiple subject headings.

For example, you might place the metadata value "water" in two subject headings--chemical or geographic. The chemical subject heading might include things like elements present (hydrogen and oxygen), physical state (solid, liquid, or gas), or chemical properties (universal solvent, etc.). The geographic subject heading might include bodies of water (lake, river, ocean), extreme weather (flood, draught), or water source (snow, aquifer, purification plant).

Note: While subject headings tend to include broader classes, taxonomies [47], the other multilevel controlled vocabulary discussed in this section of the guides, tend to include subject-specific, narrow classes.

Example

One of the most-used subject headings is managed by the Library of Congress. The LCSH (Library of Congress Subject Headings) is a vocabulary that is used to catalog printed materials available in a library setting. In the LCSH, printed media about human beings might be included in one of a variety of classes, as shown below.

Compare this to the Linnaean classification for a human being that was described in the Taxonomies guide. In that vocabulary, the *Homo sapiens* label [180] uniquely identified human beings, rather than including them in several classes.

- Library of Congress Classification Outline
- Class Q - Science
- Subclass QH: Natural history [108] - Biology
- QH301-705.5: Biology (General)
- QH359-425: Evolution
- QH426-470: Genetics
- QH471-489: Reproduction
- QH501-531: Life
- QH540-549.5: Ecology
- QH573-671: Cytology
- QH705-705.5: Economic Biology
- Subclass QM: Human anatomy
- QM1-511: General
- QM531-549: Regional anatomy
- QM550-577.8: Human and comparative histology
- QM601-695: Human embryology
- Subclass QP: Physiology
- QP1-345: General, including influence of the environment
- QP351-495: Neurophysiology and neuropsychology
- QP901-981: Experimental pharmacology
- Subclass QR: Microbiology
- QR1-74.5: General
- QR180-189.5: Immunology
- QR355-502: Virology

Notice this subject heading does not uniquely identify an object; it merely classifies it according to "like objects." The classification scheme provides us with general information about an object. Since classes in subject headings are broad, the lower objects in the lower classes do not inherit a significant amount of characteristics. For example, if something is classified with the number QH525*, we would know only that it is a scientific piece about life in natural history and biology.*

How to cite this Guide

Neiswender, C. 2011. "Subject Heading (Multi-Level Vocabulary)." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/voccat/subjectheading. Accessed June 23, 2014.

Thesaurus (Relational Vocabulary)

Definition, description and example of a thesaurus

Definition, description and example of a thesaurus

Flat Vocabularies [234]					Multi -Level Vocabularies		Relational Vocabularies [236]		
Authority File [248]	Glossary [238]	Dictionary [239]	Gazetteer [240]	Code List [241]	Taxonomy [242]	Subject Heading [243]	Thesaurus	Semantic Network [245]	Ontology [246]

Definition of a Thesaurus

In the context of metadata [5], a thesaurus is a type of relational controlled vocabulary [122] that provides a list of metadata terms with specific relationships [125] among the terms.

Discussion

According to the Guidelines for the Construction, Format and Management of Monolingual Thesauri (ANSI/NISO Z39.19 - 2003) [264], "A thesaurus is a controlled vocabulary arranged in a known order and structured so that equivalence, homographic, hierarchical, and associative relationships among values [119] are displayed clearly and identified by standardized relationship indicators that are employed reciprocally. The primary purposes of a thesaurus are (a) to facilitate retrieval of documents and (b) to achieve consistency in the indexing of written or otherwise recorded documents and other items, mainly for postcoordinate information storage and retrieval systems."

Four principal purposes are served by a thesaurus:

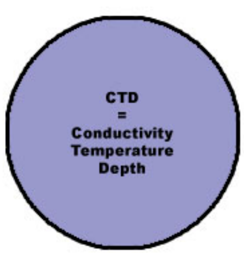
1. **Translation.** To provide a means for translating the natural language of authors, indexers and users into a controlled vocabulary used for indexing and retrieval.
2. **Consistency.** To promote consistency in the assignment of index values.
3. **Indication of Relationships.** To indicate semantic relationships among values.
4. **Retrieval.** To serve as a search aid in retrieval of documents.

There are a variety of relationships, which are described in ANSI/NISO Z39.19 – 2003 that can be used to relate terms in a thesaurus. A brief explanation of these can be found below and on page 12 of the Standard.

ABV*	Relationship Indicators	ABV*	Relationship Indicators	ABV*	Relationship Indicators	ABV*	Relationship Indicators	ABV*	Relationship Indicators
BT	Broader Term	**NT**	Narrower Term	**U**	Use	**GS**	Generic Structure	**SN**	Scope Note
BTG	Broader Term (generic)	**NTG**	Narrower Term (generic)	**UF**	Used for	**HN**	History [108] Note	**TT**	Top Term
BTI	Broader Term (instance [160])	**NTI**	Narrower Term (instance)	**UF+**	Used for... and...	**RT**	Related Term	**X**	See from (equivalent to UF); Reciprocal of *SEE*
BTP	Broader Term (partitive)	**NTP**	Narrower Term (partitive)	**USE+**	Use... and...	**SEE**	Equivalent to U (Use)		

* Abbreviation, or Thesaurus Code

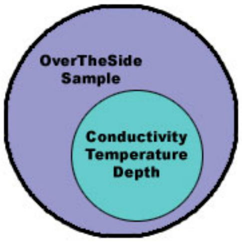
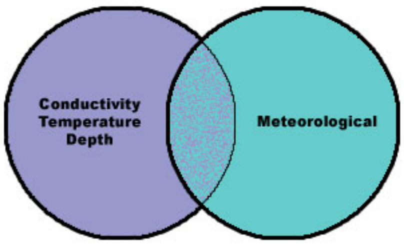
Select Abstract Examples



Equivalence Relationship
CTD USE ConductivityTemperatureDepth

USE = Equivalent

You might use the terminology "ConductivityTemperatureDepth" in your controlled vocabulary for data type. However, the scientific vernacular for this type of data is "CTD [114]". In this scenario, CTD would not appear in your vocabulary, even though many scientists' first instinct would be to search for CTD. The thesaurus could enable discovery [113] by either CTD or ConductivityTemperatureDepth.

 **Heirarchy Relationship** OverTheSideSample BT ConductivityTemperatureDepth	**BT = Broader Term** You might also use the terminology "OverTheSideSample" in your controlled vocabulary for data type. However, CTD samples are a type of over the side sampling. In this case, you could relate these two values in your vocabulary as the same variety of event. In this scenario, you could construct a query of over the side samples, and your search results would include both the more generic OverTheSideSample and ConductivityTemperatureDepth.
 **Associative Relationship** ConductivityTemperatureDepth RT Meteorological	**RT = Related To** A water temperature reading would be included both in the ConductivityTemperatureDepth data, and in Meteorological data. However, while a relationship between ConductivityTemperatureDepth and Meteorological exists, it cannot be characterized as scope, or equivalence.

Example - Alexandria Digital Library Thesaurus [265] (Selected sections of the complete thesaurus)

Alphabetical Representation

beaches
SN: The gently sloping shore that is washed by waves, usually composed of sand and pebbles. [USGS Circ 1048]
UF: sandy areas
strands
BT: physiographic features
RT: coastal zones
dunes

educational facilities
UF: academies
agricultural schools
campuses
colleges
military schools
schools
seminaries
training centers
universities
BT: institutional sites
RT: library buildings
research facilities

faults
SN: A fracture in the Earth's crust accompanied by a displacement of one side of the fracture with respect to the other. [UGSG Circ 1048]
BT: tectonic features
NT: fault zones
fracture zones
rift zones
RT: earthquake features

harbours
USE: harbors

oceans
SN: One of the major divisions of the vast expanse of salt water covering part of the earth. [NIMA]
BT: seas
NT: ocean currents
ocean regions

Comma Delimited Representation

"beaches","Used for","sandy areas"
"beaches","Used for","strands"
"beaches","Related Terms","coastal zones"
"beaches","Related Terms","dunes"
"beaches","Scope Note","The gently sloping shore that is washed by waves, usually composed of sand and pebbles. [USGS Circ 1048]"
"beaches","Broader Terms","physiographic features"
"beacons","Use","reference locations"
"educational facilities","Used for","academies"
"educational facilities","Used for","agricultural schools"
"educational facilities","Used for","campuses"
"educational facilities","Used for","colleges"
"educational facilities","Used for","military schools"
"educational facilities","Used for","schools"
"educational facilities","Used for","seminaries"
"educational facilities","Used for","training centers"
"educational facilities","Used for","universities"
"educational facilities","Related Terms","library buildings"
"educational facilities","Related Terms","research facilities"
"educational facilities","Broader Terms","institutional sites"
"faults","Narrower Terms","fault zones"
"faults","Narrower Terms","fracture zones"
"faults","Narrower Terms","rift zones"
"faults","Related Terms","earthquake features"
"faults","Scope Note","A fracture in the Earth's crust accompanied by a displacement of one side of the fracture with respect to the other. [UGSG Circ 1048]"
"faults","Broader Terms","tectonic features"
"harbours","Use","harbors"
"oceans","Narrower Terms","ocean currents"
"oceans","Narrower Terms","ocean regions"
"oceans","Scope Note","One of the major divisions of the vast expanse of salt water covering part of the earth. [NIMA]"
"oceans","Broader Terms","seas"

ADL Feature Type Thesaurus, Top Term Report, 10/03/01

hydrographic features
. aquifers
. bays
. . fjords
. channels
. deltas
. drainage basins
. estuaries
. floodplains
. gulfs
. guts
. ice masses
. . glacier features
. lakes
. seas
. . oceans
. . . ocean currents
. . . ocean regions
. streams
. . rivers
. . . bends (river)
. . . rapids
. . . waterfalls
. . springs (hydrographic)
. thermal features

regions
. biogeographic regions
. . barren lands
. . deserts
. . forests
. . . petrified forests
. . . rain forests
. . . woods
. . grasslands
. . habitats
. . jungles
. . oases
. . shrublands
. . snow regions
. . tundras

. . wetlands
. cadastral areas
. climatic regions
. coastal zones
. economic regions
. firebreaks
. land regions
. . continents
. . islands
. . . archipelagos
. . subcontinents
. map regions
. . chart regions
. . map quadrangle regions
. . UTM zones
. research areas
. . ecological research sites
. . paleontological sites

In each of the three representations of this thesaurus, the standard relationships are used to connect various values. Notice the content (values and relations) are the same in each of the representations - they are simply presented in a different format.

Additional Information

USGS Thesaurus [266], along with FAQ [267]

How to cite this Guide

Neiswender, C. 2011. "Thesaurus (Relational Vocabulary)." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/voccat/thesaurus. Accessed June 23, 2014.

Semantic Network (Relational Vocabularies)

Definition, description and example of a semantic network

Definition, description and example of a semantic network

Flat Vocabularies [234]					Multi -Level Vocabularies		Relational Vocabularies [236]		
Authority File [248]	Glossary [238]	Dictionary [239]	Gazetteer [240]	Code List [241]	Taxonomy [242]	Subject Heading [243]	Thesaurus [244]	Semantic Network	Ontology [246]

Semantic Network (Relational Vocabulary)

Definition

A type of relational controlled vocabulary [122] consisting of lists of values [119]/concepts and directed relationships [125].

Discussion

Semantic networks can be thought of as super-thesauri [49]. Each network can be represented in a directed graph of concept nodes connected by relations established in the NISO [109] standard [157], along with additional relations such as whole-part, cause-effect, or parent-child relationships. Semantic networks tend to deal more in abstractions than concrete terminology.

A finite list of relations used in semantic networks does not exist. Semantic network relations can extend to provenance [108]

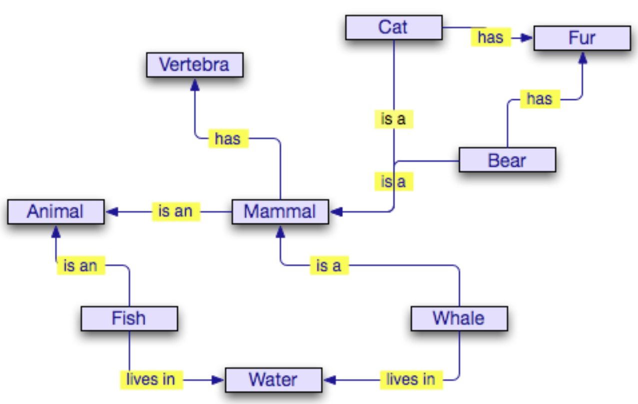
The record of how a particular value or record came to be. Provenance can include things like when, by whom, and how the item was created or modified.

(parent-child), genesis (cause-effect) and complexity (whole-part). There are also relations for opposites, antonyms, complementary concepts, possessive, passive, and others. This is in stark contrast to the clear articulation of a finite set of relations (largely based on scope), which are articulated in ANSI [233]/NISO Z39.19 - 2003 for thesauri.

Information technology experts tend to use semantic networks to establish complex search interfaces, which can help a user locate the most appropriate results based on the search term. Since semantic networks describe complex relationships, the search interface can be programmed to interpret the user entry into various nodes, which are included in a semantic network. The resulting search is more exhaustive than that provided by a multi-level set of values, because the system can be set up to return results from different levels or categories based upon relations.

Note: There is extensive research about the development and use of semantic networks in the fields of linguistics, psychology, and philosophy.

Example



This very simple diagram of a semantic network illustrates the directed nature of relationships. For example, using this diagram, you can make the statement "A fish is an animal that lives in the water." Or, "A bear is a mammal (a type of animal with a vertebra) that has fur."

How to cite this Guide

Neiswender, C. 2011. "Semantic Network (Relational Vocabularies)." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/voccat/semanticnetwork. Accessed June 23, 2014.

Ontology (Relational Vocabulary)

Definition and description of an ontology

Definition, description and example of a semantic network

Flat Vocabularies [234]					Multi -Level Vocabularies		Relational Vocabularies [236]		
Authority File [248]	Glossary [238]	Dictionary [239]	Gazetteer [240]	Code List [241]	Taxonomy [242]	Subject Heading [243]	Thesaurus [244]	Semantic Network [245]	Ontology

Definition of Ontology

A type of relational controlled vocabulary [122] that provides for categories, relationships [125], rules and axioms among metadata elements [117]. Typically a hierarchy of classes [225] and terms, an ontology is a machine-readable [112] way of relating metadata [5] terminology.

Discussion

Ontologies [107] are, by far, the most complex type of vocabulary. By definition, ontologies include multilevel vocabularies, relationships, rules and axioms. They enable the assignment of properties, comparisons using the complete list of relationships, and derived conclusions based on established rules.

For more information, please visit the Ontologies section [67].

How to cite this Guide

Neiswender, C. 2011. "Ontology (Relational Vocabulary)." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/voctypes/voccat/ontology. Accessed June 23, 2014.

Overview to Choosing and Implementing Established Controlled Vocabularies

The first steps to take in order to implement an existing controlled vocabulary

There are two ways to obtain a controlled vocabulary: start with an existing vocabulary or build your own. We strongly recommend the first approach for most situations. There are four steps in creating a system with integrated controlled vocabularies [122]. (If you have implemented your own vocabulary, you can apply steps three and four after your new controlled vocabulary is established.)

1. Choose a Controlled Vocabulary

Although just a few years ago there were very few mature controlled vocabularies for marine science concepts, today there are many possible candidates to consider. The process of finding suitable candidates, evaluating them, and deciding on the best one is described in the guide Choosing a Controlled Vocabulary [53].

2. Implement Controlled Vocabularies in Your System

The controlled vocabulary you choose must be integrated into your system. Basic implementation may involve creating and populating a drop-down menu. Because most vocabularies will change over time, keeping your system current will require some strategic decisions and planning for the future. For more information, see the guide Implementing Controlled Vocabularies in Your System [268].

3. Map Among Controlled Vocabularies

To ensure that your data will be discoverable and understandable outside your project, you may wish to map relationships [125] between relevant vocabularies. Users of some systems may need to understand which terms in your controlled vocabulary correspond to terms in a different controlled vocabulary, or they may wish to interface with your system using their own terms. In these cases, you may need to create a map between two or more different vocabularies. The guide Mapping Among Controlled Vocabularies [54] describes this process and the available tools.

4. Achieving Semantic Interoperability

Making your data and metadata [5] usable in the future, across systems and domains, is the essence of semantic interoperability. This demands use of a consistent semantic framework [269] that incorporates precision in data entry, naming of variables, and scoping of metadata descriptions so that computer systems—and ultimately the human user—can make use of your terms and your data. We address this topic in the guide Achieving Semantic Interoperability [57].

The Big Picture

Vocabularies are in their infancy. Patterns for adopting, using, and maintaining vocabularies in data systems are not uniformly followed. Mappings are relatively new and immature, and end-to-end solutions that achieve semantic interoperability are few. Nonetheless, these four steps are essential for good practices when using an established controlled vocabulary.

We encourage you to help us reference the best practices and examples in this field [117] and to help us describe the processes that provide the best results.

How to cite this Guide

Graybeal, J. 2011. "Overview to Choosing and Implementing Established Controlled Vocabularies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvchooseimplement. Accessed June 23, 2014.

Choosing a Controlled Vocabulary

So you have a content standard for your metadata, but now you need to start filling out the fields using specific terms. How do you decide whether to use a controlled vocabulary and which one(s) to use?

This guide is written for data managers and managing scientists who must implement a data system for their project. The adoption of vocabularies [130] for a metadata [5] project requires understanding of the characteristics of the project and the data system in which the vocabularies will be applied.

Finding and selecting an appropriate vocabulary takes some research. Defining your own vocabulary may seem like an easier alternative, but ultimately this approach decreases your project's ability to interoperate with other data sources.

For instance [160], combining or searching data from similar projects may require additional vocabulary mapping [217] work that could be avoided by adopting an established vocabulary.

Although the selection of a vocabulary can influence the selection of a content standard [157] so that the two can optimally work together, this guide assumes the more common situation, in which a content standard or data model has been defined as the first step. That is, it assumes the metadata fields [117] have already been defined, but the detailed terms used to fill out the fields—the vocabularies—have not been identified.

There are several factors that influence the selection of the vocabularies to use when filling out metadata fields:

- Specification [104] of a required vocabulary by the content standard
- Characteristics of existing vocabularies
    - Availability
    - Quality (completeness, clarity and precision, relevance)
    - Community adoption
- Support of content standard or data model for multiple vocabularies

This guide will help you assess whether acceptable vocabularies exist that take these factors into consideration and will help you evaluate their relative merits.

Conventions and Assumptions in this Guide

In this guide, we will refer to the document that specifies the related metadata fields as the "content standard," referring as well to any data model that serves the same purpose.

For most content standards [176], there are many different fields that require, or should require, vocabularies. Since most vocabularies cover only a single topic, usually multiple vocabularies must be selected, one (or more) for each field of the content standard. In the following discussion, we assume the context is choosing a vocabulary for a given *field* of the content standard.

Finally, it is assumed that the fields in question are best entered using terms from a vocabulary. Some textual fields are designed for ad-hoc text, and vocabularies are obviously unsuitable. Vocabularies are most suitable for those fields that have a finite number of potential terms that can be defined in advance.

The Selection Process

Assessing Your Analytic Goals

Always consider actual entries in the vocabulary to assess whether it will serve your purpose. Even if a vocabulary has very high ratings in other factors, it may not meet your project's needs. For instance, a user looking for a "sensor type" vocabulary to drive post-processing software will find most sensor type vocabularies useless because not all sensors of a given type have the same post-processing characteristics.

Conforming to the Content Standard

If your content standard specifies a vocabulary to be used, then the discovery [113] and selection process is straightforward. For example, the Directory Interchange Format (DIF) requires that the field "Parameters" be filled out from a set list of DIF science keyword categories [272]. The only concern when choosing a specified vocabulary is in using the appropriate version—either a specific version or the most recent version of the vocabulary.

Finding Available Vocabularies

If your content standard does not specify a vocabulary, then you will need to find one or more that fit your needs. There are essentially three types of vocabularies, the first of which is often the most useful:

- Those developed by a community as a general-use vocabulary.
- Those developed by a project for its own purposes, yet are useful in other contexts.
- Those that may not have been developed as a metadata vocabulary *per se*, but can be adapted to that purpose.

The first places to look for vocabularies are in catalogs, reference pages, and vocabulary or ontology servers. The Marine Metadata Interoperability [7] project (MMI [9]) provides an extensive list of vocabularies, many of which extend beyond the marine domain. MMI also provides an ontology service.

SWEET [273] is another source of earth science ontologies [107]. More general references may also serve as a source of vocabularies. In the marine domain, the IODE Ocean Portal [274] and NASA's GCMD [275] reference many resources, including vocabularies. Broader resources like Wikipedia can provide pointers to vocabularies (and can also suggest specific terms via their own entries, if you have to create your own vocabulary).

Individual projects typically have one or more vocabularies for the project, and some, like SeaDataNet [276], maintain a large number of vocabulary lists. These can usually be found by following the Data link on the project website, but a personal contact may be necessary to find or obtain the actual vocabulary. (MMI tries to represent as many of those marine and environmental vocabularies as possible and would appreciate notification of any missing from our list.)

For a particular domain or topic, a web search on "topic vocabulary" may prove useful. A number of taxonomic vocabularies (e.g., species registries) are available; see the Catalogue of Life [277] for an example list.

For science domains, like marine habitats, many vocabularies are published in individual research papers. Again, where these have come to the attention of MMI they are referenced on this site, but a literature search may uncover others.

Research libraries are also an important source of vocabularies, particularly vocabularies that have been published but not put online. Contact your institution's reference library for assistance.

Finally, word of mouth, and its online equivalent, the email forum, can still be effective sources of information. For more general vocabulary questions, the ask@marinemetadata.org [278] mail list often elicits useful information, or ask at one of the other metadata email lists pointed to by the site.

Assessing the Quality of a Vocabulary

Vocabularies can be evaluated according to criteria that are largely measurable. The relative weight of each criterion may vary according to individual needs.

Management: Is the Vocabulary Maintained Using Established and Robust Processes?

While other characteristics may be more apparent, the management of a vocabulary is the most important factor in whether the vocabulary will continue to be useful throughout the life of your project. Unless you expect the vocabulary to remain a static reference, its ability to adapt to new or changed terms will determine its long-term suitability.

Factors that reflect good management practices include a vocabulary's age, the existence and transparency of change procedures, change tracking, and publication record. More information about what to look for in these and other factors are described here.

Age: When was the last update? A vocabulary that has not been updated for more than a year is likely to be maintained slowly, if at all. Exceptions are possible if the vocabulary and domain are mature and unchanging, as can be the case for project vocabularies.

Processes: Do change procedures exist? Change procedures document how the vocabulary can be modified. Typical modifications include adding terms, improving the definition (or other characteristics) of terms, and marking or deleting terms that are obsolete. The change procedures should be clearly and publicly described. Ideally, they call for community feedback on proposed changes.

Transparency: Are procedures open and transparently followed? If changes occur without being visible in an open forum, it is difficult to be sure that they are being followed consistently and correctly. Lack of visibility also limits input from the community.

Tracking: Are changes effectively tracked? Each change made to a vocabulary should be tracked, including the date, author, original requester, and the item changed. Ideally, a reference to any related materials should be documented. Changes should be tracked at the level of individual items or records, not just at the level of whole files. Each time a change is made, the revision identifications (version number or other identifier) for any documents containing the change (e.g., the file or data set) should be updated. A single revision update may incorporate multiple item changes. Any past version of the vocabulary, or any of its terms, should be readily recoverable using either a timestamp or a revision identification.

Continuity of Presence: Has the vocabulary been consistently published? Vocabularies intended for public use should be presented in a reliably accessible online forum. The URL [279] for the most current vocabulary should not change, nor should URLs for specific vocabulary versions. All past versions should be available. Obsolete terms or definitions should remain available via archives (and not be removed from them), since previous metadata may use the obsolete terms.

Organizational Sponsorship. Although it is a relatively subjective characteristic, the nature of the organization that is maintaining a vocabulary can occasionally provide a useful clue in evaluating vocabularies. Organizations that are larger, better-funded, more permanent, and focused on good metadata practices and solutions may have an advantage here. At the same time, open source efforts that have significant community investment may have a comparably large, long-term viability, since the responsibility is spread out over many individuals, organizations, and countries.

Completeness: Is the Vocabulary Comprehensive?

A vocabulary that covers more aspects of a topic or domain is likely to be a better candidate than one providing fewer terms because it is more likely to contain usable terms. For example, a list of sensor manufacturers that only considers current commercial instrument vendors is unlikely to include vendors of all of your instruments. Such a limited list will not incorporate robust practices to distinguish between multiple phases of the same company (e.g., as company takeovers and mergers occur).

Clarity and Precision: Are Terms Intuitive, Well Described, and Unambiguous?

The ideal vocabulary completely characterizes the topic the vocabulary is designed to address. Each term is clearly distinct from every other term, and the names intuitively bring to mind the concept they represent. Descriptions for each term are sufficiently clear to eliminate any uncertainty in the user's mind about whether a term is the appropriate one.

Format: Is the Vocabulary Available Online in a Defined Format?

While many vocabularies are presented as a web page, that is, in Hypertext Markup Language (HTML), this is a difficult format to work with computationally. At a minimum, a vocabulary should be available in delimited text or Microsoft Excel format. Serious developers of controlled vocabularies [122] will present their work in an ontological language such as OWL (Ontology Language for the Web) or another RDF [187] format, so that it can be accessed online by ontological tools and downloaded for local applications.

Evaluating Community Adoption

An important consideration in choosing a vocabulary is its level of adoption in relevant communities:

- Global and regional communities
- Domain communities (e.g., research discipline or specific science domain)
- Project communities

For each community, the adoption level of a vocabulary can be assessed in non-quantitative ways. Sources of information include the vocabulary authors, managers of data systems in the community, and online searches for either references to the vocabulary or actual instances using the vocabulary.

While community adoption should not always be a dominant consideration, strong community adoption of a vocabulary can make an important difference in the value [119] of the vocabulary, especially when your goal is easier interoperability with other users' projects and systems within that community.

Advanced Semantic Relations

Vocabularies that are full-fledged ontologies, with detailed class-subclass relationships [125] and defined properties, are potentially of greater long-term value. The additional knowledge embedded in sophisticated ontologies enables using them in more advanced and more automated ways.

Supporting Tools

Some metadata editor tools have built-in vocabulary pick-lists, making implementation of those vocabularies easier. See the Tools guide for specifics.

How to cite this Guide

Graybeal, J. 2011. "Choosing a Controlled Vocabulary." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvchooseimplement/cvchoosing. Accessed June 23, 2014.

Mapping Among Controlled Vocabularies

If you need to map the terms among different vocabularies -- or you think you might need to do so, but aren't sure -- this guide provides the key information to help you decide what you need to do and how to start doing it.

Reasons for Mapping Controlled Vocabularies

Mapping provides long-term advantages that help improve the utility and longevity of a project's data.

Understanding the motivations behind mapping and the tools that are available to associate the metadata elements [117] in one project with the terms in another vocabulary will assist a metadata [5] manager in making good decisions for a project. Motivations for mapping may include the following and are described below:

- Creation of a simple term translator for terms in related communities.
- Automated translation between vocabularies created in different general standards.
- Education of users and expansion of vocabularies used in specialized communities.
- Enhancement of searchability for a project's data and metadata.
- Clarification of relationships [125] within a project's own vocabulary.
- Documentation of a project's vocabulary for posterity.
- Validation and improvement of a project's vocabulary.
- Long-term interoperability [7] with the Semantic Web [215].

Creation of a simple term translator for related communities

A vocabulary mapping can be very helpful to the user in understanding how the terms relate to each other. This situation occurs with data and related publications from different practitioners in the same domain as well as with material from different scientific areas, such as physical oceanography and atmospheric science. Having a simple presentation of the relationship among the terms in the two vocabularies can bridge this gap significantly. The mapping serves as a thesaurus [105] or small dictionary [124] for the vocabularies.

Automated translation between vocabularies created in different general standards

A more rigorous subset of the previous motivation for mapping is the need for an explicit translation of metadata records. In some cases, one metadata standard, such as ISO [128] 19115, may use one vocabulary, while another, such as FGDC's Content Standard [176] for Digital Geospatial Metadata, uses different terms for the same thing. Someone may want to have metadata in ISO 19115 for a set of data files that were documented in the CSDGM [182] standard. Creating them by hand is time consuming. A carefully built vocabulary mapping provides a tool that can automatically translate between the two languages. Note that just as in language translation, some information is likely to be lost in the process, because vocabularies rarely have perfect matches for each other's terms.

Education of users and expansion of vocabularies used in specialized communities

For practitioners in a community, the data manager can build a set of explicit relationships as a bridge from their terms to those in the rest of the science world. The mappings can enhance advertisement of their data to other communities, serve as an educational tool that will improve the entry of metadata, and expose the community's users to other vocabularies and standards. The community, in turn, can assist in verifying the mappings as they modify their own practices.

Enhancement of searchability for a project's data and metadata

Vocabulary mappings [217] make it possible to add an arbitrary number of relationships between the terms used for searches and the ones that are used to label [180] and discover the data. This enables more comprehensive and accurate search results. For example, users should be able to find similar data of interest whether they search on "sea surface temperature," "SST [121]," or "water temperature" in the same way that an online shopper might expect to find stores that sell diamonds, whether they searched on "diamonds" or "gems."

Conversely, a community may wish to provide an interface that uses formal terminology, like the Integrated Ocean Observing System's controlled vocabulary, but not incorporate the keywords from that vocabulary directly within its own data files. Mapping allows users to move easily from one set of terms to the other.

Clarification of relationships within a project's own vocabulary

A data manager may wish to document relationships between terms within the project's own vocabulary. Mapping provides a method to indicate which terms overlap with or totally include others—which are synonyms, or which are opposites. These important details assist users working with the project vocabulary and enable their systems to work more effectively with project data.

Documentation of a project's vocabulary for posterity

Mapping the terms of a vocabulary to other vocabularies provides critical context, explaining how the vocabulary fits into the larger domain. A good mapping can provide a permanent record of the meaning of many terms in the vocabulary in words that other people and computers can understand.

Validation and improvement of project vocabulary

By mapping a project's vocabulary to others, especially to a vocabulary that is more rigorous or more standardized, data managers will come to understand the strengths and weaknesses of both vocabularies involved. This kind of evaluation, paired with sharing of results can improve a project's vocabulary and help make both vocabularies part of a community knowledge-building process.

Long-term interoperability with the Semantic Web

The most far-reaching advantage of vocabulary mapping may not produce immediate value [119] for users of a particular data set or search program but may have eventual benefits in the operation of the Semantic Web in the future. The Semantic Web is a model proposed by the W3C [110] that involves the automated interaction of large numbers of computers in ways that understand the *meaning* of the words that are used. Explicit descriptions of the relationships between terms will allow the Semantic Web to provide far more effective and advanced services related to the data described by those vocabularies.

Methods for Mapping Vocabularies

There are many technologies that can be used to map vocabulary terms [209]. From simplest to most advanced, these include the following techniques:

- **Tables:** these can be in ASCII [232] files, MS Word tables, or MS Excel spreadsheets. A simple synonym table will list terms that agree, while a relationship table will include the terms from each vocabulary and the relationship between them.
- **Resource Description Framework (RDF [187]) files:** this format supports simple relationship definitions, using triples of subject, predicate or relationship, and object, to describe the connection between terms in different vocabularies. Note that all three parts of the triple correspond to a well-defined term. The relationships in the Simple Knowledge Organization System, or SKOS [210], are often used with RDF files.
- **Ontologies [107]:** these are often represented in the Web Ontology Language (OWL) format. Ontologies allow description of more complex relationships, in particular, supporting axioms about the relationships between classes [225] of individuals. Ontologies support a more formal language than SKOS.
- **Concept maps:** these are typically less formally developed than OWL or RDF relationships but also have the subject-predicate-object triple as a basis. In concept maps, which can be in the CMAP format, relationships (predicates) are often not part of a controlled vocabulary themselves. Concept maps are often used to quickly document knowledge about a domain but not to create a formal description of that knowledge that can be used for computer reasoning or inference. That is, concept maps focus not so much on relating vocabularies, as on describing the context of a domain; thus, they are not directly helpful in the type of mapping that is the subject of this guide.

For a simple mapping project, for example between two relatively structured vocabularies, an MS Excel or similar table is a manageable and, in the short term, the fastest approach (since many people are already familiar with Microsoft Office).

As a mapping project becomes more complex, tables quickly become insufficient. The problem of data entry becomes less important than the problems of finding matches, matching multiple terms simultaneously, and presenting the results in a standard format that can help verify and advance the mapping project goals.

For example, imagine a scenario in which four vocabularies have overlapping and often similar terms. The goal is to map all the terms in three of the vocabularies to a fourth, reference vocabulary, then produce a set of all the inferred relationships that result. Searching for relationships among the terms and assigning them in the map is extremely time-consuming, and obtaining fast results becomes difficult.

Preferred Technical Approach: OWL and RDF-SKOS Files

The following approaches assume that the project under consideration requires mapping and not the construction of a new vocabulary or ontology. If the mapping project will be complex or persistent (referenced for 12 months or longer), then for long-term maintainability and standardization, it should begin with either the OWL or RDF-SKOS technologies. Although each has some start-up costs to learn the technology, they both provide more advanced and reusable results and more useful tools that can quickly recoup the up-front costs.

RDF-SKOS is being adopted for many vocabulary mapping projects in which simple relationships and associations are the principle objective. The mapping relationships that have been developed in SKOS are useful for such projects, although quite complex sets of mappings among vocabularies can be developed.

On the other hand, OWL enables complex statements about many aspects of relationships between vocabulary terms and can be used in a more formal framework. It is the appropriate tool when the vocabularies themselves are already in OWL and a rich semantic framework [269] is likely to be of value in the future. OWL may also be the more appropriate tool if many different types of relationships are anticipated or if one or more vocabularies are planned to be the basis of an upper ontology.

A fairly straightforward discussion of the differences between OWL and SKOS can be found in section 1.3 of the W3C SKOS reference with additional context elsewhere in that document.

Tool Support

There are many simple and sophisticated tools for working with mappings and ontologies. For the relatively simple task of producing vocabulary mappings [219], a simple approach may be desirable. If you have a vocabulary list in the form of a text file, the first step will be to convert your list into an RDF or OWL file. The open source MMI [9] tools Voc2OWL [280] and Voc2RDF [281] can perform this transformation [218] fairly quickly.

Once your files are in RDF or OWL, the vocabulary mapping tool VINE [282]—also an open source tool from MMI—is available to perform mappings. This tool finds similar terms and easily maps them.

Among more sophisticated tools, Protégé is a free, full-featured and widely used choice and has collaboration features that are helpful in developing domain ontologies for a community. The comprehensive list of ontology tools on the MMI site [283] describes more options.

Community Challenges

In addition to technical challenges, there are cultural and attitudinal challenges associated with mapping vocabularies and caused by misconceptions in the science community:

- Lack of immediate need or value leading to unwillingness to participate
- Perception that mapping is too difficult or that processes make it needlessly complicated
- Busy PIs who are too busy to participate in mapping processes and think that the job should be done by developers
- Belief that standards and terms from other projects are not relevant to an individual PI's project

Creating effective vocabulary mappings is time-consuming and technically challenging. We strongly encourage people who need to produce vocabulary mappings to engage key members of their scientific community to participate in a focused effort.

Preferred Community Approach: Vocabulary Mapping Workshops

To produce a robust mapping, at least several domain science experts will be needed for a day or two, often on several occasions. To overcome the cultural challenges and to address the goal of producing effective mappings, community adoption of vocabulary mapping workshops provides a practical approach. The MMI project has held such workshops and developed a comprehensive template [284] for others to hold them. All are welcome to adopt this template for their own communities, although we do suggest you contact us if you plan to use it, and we request appropriate credit during your workshop. The MMI project welcomes feedback regarding the use and improvement of the template and tools.

How to cite this Guide

Graybeal, J. 2011. "Mapping Among Controlled Vocabularies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvchooseimplement/cvmap. Accessed June 23, 2014.

Using the VINE Tool

This page presents a brief guide to the MMI Vocabulary Integration Environment (VINE) tool for mapping between multiple controlled vocabularies. The guide also contains references to other sites with additional details about VINE.

About VINE

VINE is a free, open source software tool for mapping between multiple controlled vocabularies [122]. Mapping between controlled vocabularies in a given domain is essential to interoperability [7] between different data systems in the domain. In other words, it's important to have a machine-readable [112] way for a search computer to understand how terms are related (for example, if one term is equivalent to, more general than, or narrower than another term).

How to Get VINE

VINE is available as part of the MMI Ontology Registry and Repository [285]

How to Use VINE

Before using VINE, it's necessary to identify, and obtain the controlled vocabularies in the domain of interest. The best way to acquire the appropriate vocabulary files is to contact experts in the domain of interest. The files can have any extension [286], but must be in either RDF [187] or OWL [288] formats.

If you are dealing with files in other formats (text, relational data bases, Microsoft Excel), you will need to harmonize [208] them (bring them into RDF/OWL) before using VINE. The Voc2RDF tool can be used to accomplish this. The Voc2RDF [281]: online application and now integrated as part of the the MMI Ontology Registry and Repository [285]

Once your vocabulary is in RDF, there are many tools available for working with, and visualizing, the information in your vocabulary; See MMI for a list of such tools.

About Vocabulary Mapping

Doing accurate mapping requires an in-depth knowledge of the meaning of the terms in each vocabulary. It is usually most efficient to gather a small group of domain experts, who are familiar with the controlled vocabularies, together for an in-person mapping session. In addition to the domain experts, the mapping team may also include a facilitator, recorder, tools specialist, ontology specialist, domain lead, and communication liaison. (See the Vocabulary Mapping Workshop Template) [284]

In the mapping session:

1. Open VINE
2. Choose New File (this means a new mapping file)
3. Import the controlled vocabulary files
4. Begin mapping between individual terms in the vocabulary files. It may be helpful to choose one of the controlled vocabularies as a reference (or base) vocabulary to which to map other vocabularies. See Guide to the Mapping Process [289]
5. Save this .owl mapping file (and/or choose to Export it into ASCII [232] format

References

MMI VINE description [290]

Help in the VINE software

Template for Holding a Vocabulary Mapping Workshop [284], including the following two pages, in particular:

- VINE Guide [291] and the
- Mapping Process Guide [289]

How to cite this Guide

Watson, S. 2011. "Using the VINE Tool." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvchooseimplement/cvmap/cvmapvine. Accessed June 23, 2014.

Achieving Semantic Interoperability

Introduction to semantic interoperability, including definitions, core concepts, and steps toward semantic interoperability

This guide is written at an *Intermediate* level, and assumes some familiarity with most metadata [5] concepts.

Background

Semantic interoperability [292] exists when different systems - in our case, computer systems, interacting with each other or with people - can interact and make effective use of the terms that are used in the interaction. For example, a meteorology model may use 'temperature' for air temperature, and provide a more complete term for 'water temperature'. This may work fine with another meteorological system, but will cause confusion when interacting with an ocean model where 'temperature' means water temperature and 'air temperature' is the alternative.

The same concept exists in human interactions, with the most obvious lack of semantic interoperability occurring when people speak different languages. The key difference in creating semantic interoperability for people and computers is the amount of precision that is needed.

When two people try to communicate, a lot of redundant information is available for confirming assumptions and refining understanding. Facial expressions, tone of voice, repetition using different words, gestures, actions, and physical objects themselves guide the participants' understanding. Also, many terms can be (and are) approximately translated - for many purposes, 'friend' and 'colleague' can substitute for each other.

However, computers typically do not transmit redundant information, do require precise correspondence of terms, and have an extremely limited set of communication protocols [132] to fall back on when terms do not produce expected results. Thus, creating semantic interoperability among computer systems requires significantly more attention to detail than creating it among people.

Just as related businesses must achieve semantic interoperability across all their computer systems, scientific software increasingly demands semantic interoperability. As data and metadata go from device manufacturers, into observatories, through local repositories, post-processing algorithms, and national archives and clearinghouses, confusion about the meaning of the terms will make the information increasingly difficult to find or use. The passage of time between the original experiment, and the analysis that increasingly often take place years or decades later, only increases the loss of usability.

Core Concepts

Semantic interoperability may never be perfectly seamless and automatic, but with proper data stewardship it can be nearly so for most systems. At a minimum, the originators of data can make sure their data will remain usable by scientists and educators for many decades, and across all science disciplines.

For people speaking different languages, dictionaries [124], phrase books, and translation systems are used to provide some minimal level of semantic interoperability (i.e., communication). Similar concepts apply in computer science, with metadata dictionaries (controlled vocabularies [122]), ontologies [107], and a standards-based semantic framework [269] corresponding to the human-oriented tools. This guide describes how these pieces fit together, and what steps you will need to take to ensure your data fits into the puzzle successfully.

First, a content standard [176] or specification provides a format and specification to describe data and metadata. Ideally the specification requires the structure of the data to be fully and precisely described, so that computers can automatically parse the data into its original components. A good content standard will define all the fields and terms that it uses, making clear whether 'Data Originator' means the principal investigator, the device operator, the institution paying for the experiment, or the device itself.

As a particular description is filled out according to the specification, individual elements must be filled in with data or words. Again, where an element is filled in with specific text like keywords or codes, good specifications [104] describe what terms can be used to fill in the element - the 'controlled vocabulary' or dictionary to use when filling out that element. (Alternatively, the specification may tell the preparer how to specify which vocabulary was used to fill in the element.) In any case, a computer should be able to automatically look up each term and its meaning if it knows how to interpret descriptions that follow the specification.

But if all the computer can find is a free-text definition for the term - for example, looking up "tropo" gives the definition "tropospheric region of the earth's atmosphere" - it may be impossible for the computer to recognize that 'tropo' in this data set is the same thing as 'troposphere' in another data set. A mapping must exist between the vocabulary terms [209] that are used to describe the data set, and other vocabulary terms used to describe other data sets. With such a mapping of terms, the computer can make informed judgments about all the controlled vocabulary fields in the metadata, such as quality control flags, units, science domains, and topic keywords. Building on the first two components, this ability to connect concepts across data sets and data systems is what finally creates semantic interoperability.

A semantic framework provides an infrastructure that can use these mappings, and the information associated with them, to solve real-world science data management issues. A consistent semantic framework will include specifications for how to refer to a specific term from a specific vocabulary; how to create and understand mappings from one specific term to one or more other terms, possibly in another vocabulary; and how to build software that uses these services to give the user what he or she really wants. So if the user types in 'troposphere' as a search term to a web interface, the developer who uses a good semantic framework can build in the software tools that translate that term to all the other terms that have been mapped to it, and so the web interface will return resources to the user that it could never have found without the semantic framework.

Steps Toward Semantic Interoperability

The first steps of semantic interoperability depends on a foundation of good data practice, principally defining (in a standard way) the way your data and metadata are structured. The time to do this is when you create the data, because that is when you best understand how the data are organized. Other guides in this series describe this process.

Once your data and metadata structures have been defined - typically by using a content standard to organize your metadata, and metadata to describe your data's structure - then you can focus on describing the data in a semantically interoperable way. Where the structural information might say you have 3 variables in ASCII [232] format separated by tabs, semantic interoperability demands that you name those variables, and that some correspondence exist between your variable names and the names that other people - and computers - recognize.

Making your names understandable can take several forms. The easiest in some situations is to choose a vocabulary that can describe all of your variables - for example, the COARDS Climate and Forecast Standard Variables [293] is one with extensive coverage - and declare that all your names will be specified using that vocabulary.

Other options are available for other situations. Need to specify terms from multiple vocabularies? You can specify the names using a syntax that includes the vocabulary name. Need to use your own local names? You can specify a relationship [125], or vocabulary mapping [217], that people can use to relate your names to more common terms in another vocabulary. You need to use multiple vocabularies? Such mappings can connect your terms to many different vocabularies, by using a standard framework that can reference the vocabulary and the term. Even if no term from another vocabulary fits exactly, you can describe the relationship between your term and one in another vocabulary (e.g., my term is "narrower than" the other term).

Not all of these solutions are totally in place for every scenario, but most of the pieces are in place, and groups like MMI [9] can help guide you in more advanced cases. As more projects need semantic interoperability, and implement the approaches that have been created, you can be sure that your investment will either produce fairly widespread interoperability, or make it trivial to do so as missing pieces of the infrastructure develop. In most cases, the initial cost will be only a small increment above the work you will be doing anyway, and the return on your investment will be evident early and often.

How to cite this Guide

Graybeal, J. 2009. "Achieving Semantic Interoperability." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvchooseimplement/cvsemint. Accessed June 23, 2014.

A Last Resort: Developing a Local Vocabulary

Developing your own controlled vocabulary

Controlled Vocabulary Management

First, you must choose whether to use existing controlled vocabularies [122], or to implement and manage your own vocabulary. Management tasks can be avoided if you use the vocabularies managed by another organization. This is usually a good idea, as it will save time and effort and maximize sharing of terminology (see Choosing and Implementing a Controlled Vocabulary [294]).

However, a controlled vocabulary may not exist that meets your project's needs. In that case, your group will need to create and manage a controlled vocabulary (see A Last Resort... Developing a Local Controlled Vocabulary [295]).

Regardless of whether you choose to use an established controlled vocabulary or create your own, it is useful to understand the processes of developing and managing a controlled vocabulary:

Simplified Controlled Vocabulary Development and Management

1. Clearly define the need for a new controlled vocabulary and determine its specific requirements. Individuals or groups that manage controlled vocabularies must meet the needs of the relevant scientific and technical communities.
2. Using community expertise, evaluate each candidate term. Is the term widely used? Does it have appropriate meaning to the community?
3. After a thorough review, format the controlled vocabulary. Different types of controlled vocabularies can be implemented using different formats.
4. Register the controlled vocabulary with an appropriate organization.
5. Use the controlled vocabulary in community projects. Solicit input from implementing organizations.
6. Incorporate user community input to improve future versions of the controlled vocabulary.

Evolution of Controlled Vocabularies

An organization could begin with an authority file [223], then provide descriptions and etymology in future versions of the controlled vocabulary. This will enhance the authority file and transform it into a dictionary [124]. Perhaps one of the implementing organizations will enrich the dictionary by submitting classifications, relationships [125], and axioms to the managing organization for the dictionary. What started as an authority file has now become an ontology [107]/dictionary combination. (These terms are explained in Classification of Vocabularies.)

The controlled vocabulary may evolve through contributions by implementing organizations, and can become a living resource that is relatively easy to update, enhance and understand.

It is preferable to use existing standards for long-term interoperability [7]. However, if existing vocabularies are not sufficient, even with extensions, it may be necessary to create a customized controlled vocabulary, whether for new projects or for legacy data.

Considerations for Creating a Vocabulary

These considerations are valid for new projects as well as for legacy data and are explained in more detail below.

- Identification of all terms and values [119] as discrete content
- Separation of embedded information
- Clarity of units
- Inclusion of natural terms
- Reduction of ambiguity in definitions
- Consistent syntactic rules
- Grouping of terms for discovery
- Scalability [296]
- Allowing for user input
- Identification of discrete content

Identification of all terms and values as discrete content

The starting point is creating a list of all terms and possible values. To identify the terms of the vocabulary, you need to first examine the descriptions of your assets, looking for discrete (that is, non-continuous) content. Things that are measured are usually continuous, that is, they may have a limitless number of values. Terms whose values have specific descriptions are usually discrete, and any term for which the total number of possible descriptions can be counted is likely to be discrete.

If the possible content of the metadata element [117] is found to be discrete, then it is a likely candidate for a vocabulary. For example, if the descriptor is ocean_name, and the content is the name of the ocean, then the five ocean names could be added to the system as terms in a vocabulary.

Once you have identified those elements that contain discrete terms, you must identify all possible terms to be contained in the elements as values. This is the list of terms for the vocabulary. A definition of each value should exist, such that its definition is unique to that value. This definition development is a process of building a dictionary of values for the vocabulary.

Separation of embedded information

Vocabulary terms [209] should not include embedded information in the values. A value that contains encoded information may have certain characters that include facts about the value without any explanation. For example, a single value like "XT07aa" might indicate an XBT [297] temperature from a T-7 computed using coefficient set aa. This example value contains information on the type of sensor, the model of sensor, the parameter being measured and processing information. Each of these pieces of information should be split out of the single value, into separate terms and values.

Clarity of units

Units are important. Your usage vocabulary [123] may or may not contain explicit units. For example, the data values in the usage vocabulary may have a direct association with the unit (that is, one term can only have one unit). A preferred method is to allow multiple units for a single data value (for example, distance can have units of meters or kilometers). By allowing multiple units you effectively introduce another type of vocabulary that your system must support—a unit vocabulary.

Inclusion of natural terms

Whenever possible, natural terms that are commonly used within the community should be used in the vocabulary.

Reduction of ambiguity in definitions

This consideration is the counterpart of inclusion of natural terms. If terms introduce ambiguity, then consider other terms. The terms used in your vocabulary should be associated with rigorous definitions and these definitions should be unambiguous to the community using the vocabulary.

Consistent syntactic rules

The terms used in the vocabulary will be created using a set of syntactic rules that may involve capitalization, the use of underscores, or other special characters. The vocabulary must be developed with consistent application of these rules.

Grouping of terms for discovery

Values that are associated with the terms in the usage vocabulary may be grouped, effectively creating a discovery vocabulary [113]. Allowing for such grouping will help in the management of both vocabularies and the discovery of terms by users. The vocabulary should be capable of accommodating this grouping with minimal impact on the management system.

Scalability

Allowing for additions to a vocabulary is an important aspect of planning. The vocabulary should not be limited by the initial terms and values in the list. To avoid this, the term list must consider the general class of things that each term describes and allow for attributes to be defined beyond the immediate terms. For example, if you were studying highway traffic and defined the acceptable values for the Number of Doors term, you might accept 2, 3, 4, or 5 as acceptable values for the number of models of cars. However, you may wish to broaden the term to its more general description as a vehicle and add 0 as an acceptable value for Number of Doors to allow for motorcycles and scooters.

The process of defining attributes of general classes [225] is a good step towards developing an ontology, which is discussed in a subsequent guide.

Allowing for User Input

Users need a mechanism to suggest new terms for the vocabulary without giving them the direct ability to add new terms. A vocabulary is controlled to avoid confusion among terms and to avoid the introduction of errors. Additions, deletions or corrections must be managed by the person responsible for the vocabulary.

Special considerations for legacy data

Creating a vocabulary for data after a project has ended requires the same considerations as for a new project. In addition, the process may require additional effort, since only archival information is available to define terms. The data custodian may have to act as data system designer, scientist, and detective to obtain all the information necessary to create an interoperable vocabulary.

The topic of creating controlled vocabularies for legacy data is explored in more detail in the guides that follow.

How to cite this Guide

Isenor, A. 2011. "A Last Resort: Developing a Local Vocabulary." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvdev. Accessed June 23, 2014.

Developing Controlled Vocabularies

Information about creating a new vocabulary

The following should help you define the terms in your new vocabulary. Remember, you have to make the vocabulary expandable (scalable [296]) because there will likely be additions. As well, there are a few tricks you can think about before starting.

Have a specific question about developing vocabularies [298]? Ask MMI! [278]

How to cite this Guide

Isenor, A. 2009. "Developing Controlled Vocabularies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvdev/cvdevnew. Accessed June 23, 2014.

How to Determine the Terms

Description of the selection of metadata values to include in a vocabulary

To identify the terms of the vocabulary [130], you need to first examine the descriptions of your assets, looking for discrete (i.e., non continuous) content. Things that are measured are usually continuous, while things that have specific descriptions are usually discrete. Also, if you can count the total number of possible descriptions, it is likely to be discrete.

If the possible content of the metadata [5] element [117] is found to be discrete, then it is a likely candidate for a vocabulary. For example, if the descriptor was ocean_name and the content was the name of the ocean, then the ocean names could be added to the system as terms in a vocabulary. In this case, the vocabulary contains the five ocean names.

Once you have identified those elements that contain discrete terms, you must identify all possible terms to be contained in the elements as values [119]. This is the list of terms for your vocabulary. You should be able to provide a definition of each value, such that its definition is unique to that value. This definition development is a process of building a dictionary [124] of values for the vocabulary.

Have a specific question about vocabulary values? Ask MMI! [299]

How to cite this Guide

Isenor, A. 2009. "How to Determine the Terms." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org /guides/vocabs/cvdev/cvdevnew/cvdevnewterms. Accessed June 23, 2014.

How to Create a Scalable Controlled Vocabulary - Allowing for Additions

The process of creating a scalable vocabulary by allowing for the addition of values

The scalability of a vocabulary [130] is an important aspect. The vocabulary should not be limited by the initial terms in the list. To avoid this, you need to examine the terms and think about the general class of things that all the terms are describing. Don't think about an individual term (or an individual car, to extend the vehicle example [40]). Rather, think about the general class of things. Now, attempt to define attributes of the general class. This may not be an easy process. However, if you are successful your vocabulary will be scalable [296]. This process is also an excellent step towards the development of an ontology [107].

Have a specific question about vocabulary scalability? Ask MMI! [299]

How to cite this Guide

Isenor, A. 2009. "How to Create a Scalable Controlled Vocabulary - Allowing for Additions." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvdev/cvdevnew/cvdevnewscalable. Accessed June 23, 2014.

Tips and Tricks

A few tips to use in developing a vocabulary

Don’t Have Vocabulary Terms with Embedded Information

Don’t encode information within the vocabulary values [119]. As an example, a value that contains encoded information may have certain characters as meaning certain facts about the value. For example, a single value like XT07aa might indicate an XBT [297] temperature from a T-7 computed using coefficient set aa. Such a value contains information on the type of sensor, the model of sensor, the parameter [117] being measured and processing information. This type of information should be split out of the single value, into multiple values.

Think About Future Grouping of Terms

At some point, you may have to start grouping values associated with the terms in the usage vocabulary [123]; effectively creating a discovery vocabulary [113]. Allowing for such grouping will help in the management of both vocabularies [130] and in the user discovery of terms. Your vocabulary management should be capable of adding this grouping with minimal impact on the management system.

Don’t Allow Users to Manage the Vocabulary

Users need a mechanism to suggest new terms for the vocabulary but they cannot be given the ability to add new terms. A vocabulary is controlled to avoid confusion among terms and to avoid the introduction of errors. Additions, deletions or corrections must be managed by the person responsible for the vocabulary.

Units are Important

Your usage vocabulary may or may not contain explicit units. For example, the data values in the usage vocabulary may have a direct association with the unit (i.e., one term can only have one unit). A preferred method is to allow multiple units for a single data value (e.g., distance can have units of meters or kilometers). By allowing multiple units you effectively introduce another type of vocabulary that your system must support—a unit vocabulary.

The Same Syntactic Rules

The terms used in the vocabulary will be created using a set of syntactic rules that may involve capitalization, the use of underscores, or the use of other special characters. The vocabulary must be developed with consistent application of these rules throughout the vocabulary terms.

Use Natural Terms

Whenever possible, natural terms that are commonly used within the community, should be used within the vocabulary. However, if these terms introduce ambiguity, then consider other terms. Unambiguous terms and definitions are the cornerstone of the vocabulary.

Unambiguous Definition

The terms used in your vocabulary should be associated with rigorous definitions. These definitions should be unambiguous to the community using the vocabulary.

Have a specific question about developing vocabularies? Ask MMI! [299]

How to cite this Guide

Isenor, A. 2009. "Tips and Tricks." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvdev/cvdevnew/cvdevtips. Accessed June 23, 2014.

Developing Vocabularies for Legacy Data

Information about creating a vocabulary within an existing metadata project

Creating a controlled vocabulary [122] for data that already exist requires commitment to a long-term plan. The data manager must apply the considerations outlined in previous guides about developing local vocabularies [130], as well as take into account that some information about data collection may not be available after the project is complete. The data manager creating the vocabulary may not have been part of the original project and needs to consider the following:

Knowledge of the Field

Developing vocabularies requires knowledge of the field from which the data originated. Knowledge of the field will provide the data manager with the ability and credibility to talk to the people who really know the data–the scientists who collected or produced it.

Existence of Archived Reports or Planning Documentation

For legacy data, valuable sources of information may exist in documents produced during data collection, processing, or the reporting of results. These documents should be searched for metadata [228] relevant to the data set, keeping in mind that the planning documents for data collection may differ from the actual collection.

Division of the Data into Subsets

Dividing the </ could be further subdivided into scientific topics, such as physics, chemistry, biology, or geology. Choosing the chemistry topic of one cruise would be a suitable starting place in this example, if that were the area best known by the data manager creating the vocabulary.

Examination of the Data

Data can be categorized by instrumentation, procedures, and units. Compiling lists of data types and their corresponding unit names, along with instrumentation, collection procedures, and processing procedures will form the starting point for creating local vocabularies. All allowed units must be included.

Usage vocabulary

Creating a usage vocabulary will require investigation of the provenance [300], or history [108]

The record of how a particular value or record came to be. Provenance can include things like when, by whom, and how the item was created or modified.

, of the data names and values [301] associated with these names.

Confirmation of data names and values

Careful examination of data names and values will include identification of when data with different names may actually be the same, or when data with the same name may be different.

As terminology is compared, identified, and changed, this process should be documented, since it will be extremely useful in the development of a thesaurus [105], metadata mappings, and general documentation.

Clarity of units

A considerable amount of complexity exists in the domain of units. If units are abbreviated differently, then they are different (for example, oxygen content in mg/l is not equivalent to ml/l, even if the values seem similar).

Differentiation of procedures

Research settings (for example, a chemistry or biology lab or an instrumentation development shop), are full of cases where multiple procedures exist to measure the same data type. If two different procedures or instruments measure the same parameter [117], then the name associated with the data should be the same in the usage vocabulary.

A procedure vocabulary will record the differences in the measurement procedures.

Procedure or Instrumentation Vocabularies

When different procedures quantify the same data type, the information detailing the different procedures must be documented and maintained with the data value in a separate, instrumentation/procedure vocabulary for each data type.

How to cite this Guide

Isenor, A. 2011. "Developing Vocabularies for Legacy Data." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvdev/cvdevlegacy. Accessed June 23, 2014.

How to Approach Developing Vocabulary for Legacy Data

Guidance for creating and managing a vocabulary for legacy data

It is important that you approach this task with a long term vision and commitment to that vision. To do this task properly, will first and foremost require knowledge in the field from which the data originate. Knowledge in the field will provide you with the ability and credibility to talk to the people who really know the data – the scientists who collected or produced it. This knowledge will require time to build.

The first thing you will need is a plan for vocabulary development. This plan may actually be a subcomponent of a larger plan – a plan for metadata [5] capture and management. However, here we only deal with the vocabulary subcomponent.

You should start with a small subset of the data. Divide the entire data set that you will be dealing with into logical pieces (logical from your point of view). If your data asset is a collection of many data sets collected from oceanographic cruises, start with a single cruise. Then subdivide further, perhaps by topic (physics, chemistry, biology, or geology). Start with the topic you know the most about.

Now examine the data. You should be looking for the different types of data that were collected, the different instrumentation or procedures that were used and different units that may be possible for the data types. You can start with compiling a list of names that refer to the data types. Also make a list of allowed units for those names. Finally, start to document the procedures followed to collect or process the data type (if you are lucky, there will be existing documentation on procedures). These lists will form the basis of your vocabularies [130]. For example, the list of data types will form the starting point for your usage vocabulary [123].

The usage vocabulary will require a bit of extra work on your part. You should investigate the provenance or history [108]

The record of how a particular value or record came to be. Provenance can include things like when, by whom, and how the item was created or modified.

of the data names and values [119] associated with these names. During this process you should examine the various data quantities and the names affixed to these quantities. Ask yourself if two quantities with different names are actually the same, or alternately, if two with the same name are actually different. This terminology evolution should be documented, as it will be extremely useful in the development of a thesaurus [105], metadata mappings, and general documentation. You should think about different procedures for acquiring or processing the data. Finally, don't forget units and don't underestimate units. A considerable amount of complexity exists in the domain of units – and if the units are abbreviated differently, they are different (e.g., don't think for a second that oxygen content in mg/l is equivalent to ml/l; even if the values are similar).

In this process, no detail is too small. The research environment is full of cases where multiple procedures exist to measure the same data type. For example, two different biological incubation setups may produce measurements of the same data quantity. These different procedures represent important metadata that needs to be associated with the data quantity. However, the usage vocabulary needs to indicate the same data term is being measured. Another vocabulary notes the differences in the measurement procedures.

At this point you should start to realize that your job as data custodian has been morphed into a combination of data system designer, scientist, investigative police officer and investigative news reporter.

Have a specific question about vocabularies? Ask MMI! [278]

How to cite this Guide

Isenor, A. 2009. "How to Approach Developing Vocabulary for Legacy Data." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvdev/cvdevlegacy/cvdevlegacyapproach. Accessed June 23, 2014.

Tips and Tricks

Various suggestions for developing vocabularies for legacy data

The “Dos and Don’ts” from the previous section [62] also apply to legacy data. The following apply as well:

Archived reports are valuable

For legacy data, there will likely be a limited source of information in documents produced during data collection, processing, or the reporting of results. These documents are valuable resources and will need to be searched for metadata [5] relevant to the data set. However, if you find planning documents that outline planned data collection, keep in mind that these represent the planned collection – actual collection will likely be different.

Different procedures that quantify the same data type

In a research setting (e.g., a chemistry or biology lab, or instrumentation development shop), you should be looking for different procedures that quantify the same data type. For example, procedure A and procedure B (or instrument A and instrument B) may both be used to measure values [119] for a single data type. Both procedures (or instruments) measure the same parameter [117], and as such, the name associated with the data from both procedures should be the same.

However, the information detailing the different procedures must be documented and maintained with the data value. Thus, a separate vocabulary detailing the procedures should be created. This vocabulary is effectively an instrumentation/procedure vocabulary that notes for each data type the procedures used to determine the data value.

Have a specific question about vocabularies [130]? Ask MMI! [299]

How to cite this Guide

Isenor, A. 2009. "Tips and Tricks." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides /vocabs/cvdev/cvdevlegacy/cvdevlegacytips. Accessed June 23, 2014.

Vocabulary for Legacy Data: Data Rescue

Sample description of the development of a vocabulary for legacy data

Legacy data can be found in many formats and in varying condition. A data rescue project is an effort toward recovering data that are presently inaccessible.

In the hypothetical case below, we are rescuing data that were collected by multiple scientists over more than thirty-year careers, much of it before common computer use. The data being rescued may be in paper form in filing cabinets stored in the basements of office buildings or similar locations.

The data represent a wealth of historic information that are irreplaceable and are now in jeopardy of being lost. The data may contribute to long-term data sets, a particularly important topic for understanding global trends. These data need to be rescued and placed in managed databases at the organizational and national levels.

Initial Investigation and Data Evaluation

Applying the recommendations from previous guides, the first step is to begin collecting information from the scientists. They, or possibly those involved in the field programs, may have documentation on data collection plans that pertain to individual data sets. There may be logbooks or field journals that were used for notes during the field activity. Reports may exist that describe the actual field programs that resulted in the data sets. The data rescue project will need to undertake each of the steps below for a full understanding of the data being rescued:

- Collect examples of the data, paper listings, or plots.
- Identify the types of collected data; notes taken during this step should include the sampling procedures or instruments used for the particular data.
- Keep notes on when the data were collected and a general idea as to where they were collected; spatial information may be useful for prioritizing the rescue effort (for oceanographic data, Marsden Squares [302] may be useful to approximately locate data samplings).
- Determine if there are hardcopy or electronic records (a particularly difficult problem is when the data exists as a physical sample).
- Discover any backup copies.
- Determine if there is any other activity currently underway to rescue these data.

The terminology used for the collected data will be the starting point for the usage vocabulary. However, when initially scanning the documentation and legacy data, the emphasis should be on building knowledge about the collected data, rather than on building a vocabulary.

Review, Clarification, and Reduction of Terms

After reviewing numerous data sets from various scientists, the next step will be reviewing the notes from the first step about the data that were collected. When comparing the input from the scientists, there is likely to be different, but similar, terminology used for the collected data. The scientists may have used similar procedures or instruments to collect data but named the data differently. Based on the notes and comparisons of findings, revisiting the scientists will help to clarify if the data names that were noted as different, are in fact the same (or, if the same names are really different elements).

This process will likely reduce the number of terms in the list of element types, since different terms will refer to the same element.

Creation of Terms and Value Lists

With this refined list, the next step is defining the other important attributes for the terms, such as the date of formal creation of the term and the limits on the values [303] associated with the term. If the information will be stored in a database, each term will need a unique identifier, as well as a short description of the term and a longer, more detailed, description as to what the term means. This list of terms forms the usage vocabulary [304] for the project.

Comparison with Organizational and National Terminology

If the organizational and national terminology meets the needs of the legacy data rescue project, then the project should use the organizational and national usage vocabularies [123]. In that case, this stage is the proper time to identify and suggest updates to the established vocabularies. If the existing vocabularies are not adequate for the rescue project, then the project must create its own local vocabulary based on the list of terms identified above.

Long-Term Management of the Vocabulary

If the data rescue project results in a new, local controlled vocabulary [122], it will need to be managed. The data manager who established the vocabulary is likely the person with the most knowledge about this vocabulary and project and therefore the best person to be responsible for its management.

How to cite this Guide

Isenor, A. 2011. "Vocabulary for Legacy Data: Data Rescue." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/cvdev/cvdevlegacy/cvdevlegacyexample. Accessed June 23, 2014.

Ontologies: Relational Vocabularies

Data managers, scientists, and others in the marine science community are used to working with controlled vocabularies [122], even if they know them by another name. After all, anyone working with a set of data described using an agreed-upon set of terms has been exposed to a controlled vocabulary.

Controlled vocabularies are a good tool: they are often simple to develop, can be used among a small community of users, and are easy to store, visualize, and access.

However, as projects and users require more data interoperability [7], more advanced data comparison, and better discovery [113], the limitations of simple vocabularies become more apparent. Ontologies [107] provide additional functionality to meet these challenges.

Ontologies improve on simple vocabularies by allowing relationships [125] between terms to be defined (e.g., "Sea Surface Temperature" is a kind of "Temperature"), and by providing a way for rules or properties to be defined for terms (e.g., an "estuary" must have a salinity of <33 psu).

Semantic technology [305], including the use of ontologies, paves the way for data interoperability, advanced search and discovery, and "machine reasoning" in a way that simpler technologies cannot support.

In this guide, we will explore the nature of ontologies and their applicability in marine science. First, we will explain exactly what an ontology is, including how it differs from a standard [157] controlled vocabulary. Then we discuss the importance of ontologies, including their various strengths. We also provide a brief overview of the technologies which form the foundation of ontologies, including the RDF [187] and OWL [288] formats. Finally, we discuss various methods of developing, providing, and working with ontologies.

How to cite this Guide

Alexander, P. 2011. "Ontologies: Relational Vocabularies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont. Accessed June 23, 2014.

What is an Ontology?

Definition of an ontology

An ontology is a representation of knowledge, generally of a particular subject (domain), written with a standardized, structured syntax. An ontology contains concepts (resources), which serve to characterize the domain.

An ontology can relate resources to other resources, either internally or in other ontologies [107]. Resources can represent the existence of an individual entity ("grannySmithApple445" or "cellPhoneOwnedByJerry"), define subclasses that have a relationship to a class ("grannySmithApple" subclass of "apple", "cellPhone" subclass of "telephone"), and define a class to be associated with a group of entities or subclasses ("fruit," "communication device"). Resources can be associated and defined using relationships [125]. For example, an individual resource is associated with a class ("apple" is a member of "fruit") or a class is associated with an ontology (class "fruit" is described in an ontology called "food").

Ontologies vs. Controlled Vocabularies

A formal specification [104] of a vocabulary can be something we are all familiar with: a plain list of words, a dictionary [124], a taxonomy, or a thesaurus [105]. Or, it can be a more technical document: an Entity-Relational (ER) diagram, an Object Model in Unified Modeling Language (UML [306]) diagram, or an eXtensible Markup Language (XML [131]) schema [185]. Many other representations are possible for controlled vocabularies.

There are two views on what makes a controlled vocabulary qualify as an ontology. In the first view, simply expressing the vocabulary in an OWL file makes it an ontology, and further subtleties of classification are not important.

In the second view, a controlled vocabulary becomes an ontology when its concepts are defined explicitly and at least some of them are defined as classes. In addition to this requirement, an ontology needs to conform to strict hierarchical subclass relationships between the classes [Gruber, 1993]. The trivial ontologies that simply specify some terms that do not serve any other purpose than naming, to many ontologists do not further the semantic web [215]. Their lack of classes, relations, and properties makes them insufficiently powerful to be designated as ontologies.

Ontologies can include all of the following, but are not required to include them, depending on which perspective from above you adhere to:

- Classes (general things, types of things)
- Instances (individual things)
- Relationships among things
- Properties of things
- Functions, processes, constraints, and rules relating to things

Other Considerations

It is worth noting that a science, also called Ontology (denoted here with a capital "O"), has existed for several centuries and has helped to inform current practice in the computer and information sciences, though major differences do exist. Ontology is the study and description of reality, or what can be said to exist, and an attempt to categorize existing things and their relationship to one another. While Ontology seeks to describe every possible thing, ontologists in computer science tend to work in particular knowledge domains, focusing their work on smaller portions of a larger ontological whole.

Some of these terms, particularly 'ontology', have been defined many different ways in different publications. Deborah McGuinness, for example, has proposed that an ontology could be construed as including the entire spectrum of controlled vocabularies. In this guide we use one of the more common definitions, but usage in other papers or contexts may vary. In The Semantic Web, the authors refer to the 'ontology spectrum,' ranging from weak semantic entities like taxonomies [224], to strong semantic solutions like conceptual models and advanced logics.

References:

- T. Gruber, A Translation Approach to Portable Ontology Specification., Knowledge Acquisition 5(2), 199-220, 1993.
- M. C. Daconta, L. J. Obrst, K. T. Smith, The Semantic Web. Wiley Publishing © 2003. (pp 156-158 ff).

How to cite this Guide

Alexander, P., Graybeal, J. 2011. "What is an Ontology?." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/definition. Accessed June 23, 2014.

The Importance of Ontologies

We've discussed controlled vocabularies [122], what they are, how they are created, and how they are put into use. Using controlled vocabularies in the form of ontologies [107] gives you all the benefits of normal vocabularies with the added robustness of ontologies. Though there is a learning curve, when you begin to see vocabularies from an ontological point of view, the potential for more complex relationships [125] in your data, greater interoperability [7], and using computer reasoning becomes apparent.

We are just beginning to see the potential of ontologies, especially in the realm of the semantic web [215]. The Open IOOS Sensor Web demonstration [307] shows a small piece of what ontologies can do for the marine sciences. Using an OGC Sensor Web [308], Open IOOS has been successful in creating a real-time sensor map that provides data from over 1000 observation platforms. While the demonstration is a significant step, future projects, such as the Ocean Observing Initiative [309] Cyberinfrastructure project [310] will put entire worldwide systems into operation that use a semantic infrastructure as the basis for much of their data framework.

Looking beyond projects such as these, it is conceivable that data from marine science could interoperate with all of the sciences, and vice versa, enabling powerful insights into large-scale phenomena. By laying the infrastucture now, we will be prepared to participate in and lead these projects as they become reality.

How to cite this Guide

Alexander, P. 2011. "The Importance of Ontologies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/importance. Accessed June 23, 2014.

Document Domain Knowledge

From many perspectives, the goal when creating an ontology is to effectively describe a particular domain or "universe." Given the number of assumptions that go into even a single individual's understanding of their discipline or their area of research, capturing all the relevant knowledge can seem overwhelming. However, ontologies [107] force communities to think about these assumptions in very explicit terms, as they capture and describe the data and relationships [125] in the ontologies they will use in their work.

Because ontologies use a formal grammar and structure, we can tell computers how to analyze the information they contain. This structure and the myriad of rules employed in ontologies, allows computers to analyze concepts and data in new ways. By following the structure, humans can describe the world in a way that computers can understand. As a particular community creates more descriptions using ontologies, the model of their domain becomes ever more complete, and the power of the collected ontologies grows.

Considering that ontologies are built from resources and the relationships between those resources, it may seem surprising that the knowledge of diverse communities can fit into the single structure. In fact, different communities may need to define new relationships to describe how their resources are related. The formal grammar of ontologies lets each community describe the nature of those custom relationships—for example, specifying whether a relationship is transitive, or whether it is the inverse of another relationship—so that computers can perform operations using them.

The final and perhaps most powerful element [117] of working with domain knowledge involves inferencing, a technique that allows computers to create new relationships and meaning based on what they already know through previous ontologies. In other words, computers can discover information that a human didn't explicitly describe for them. This provides exciting opportunities both for communities that produce ontologies, and for other communities that can make use of the ontologies.

How to cite this Guide

Alexander, P. 2011. "Document Domain Knowledge." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/importance/domainknowledge. Accessed June 23, 2014.

Document and Develop Community Understanding

Besides documenting a domain of knowledge, ontologies [107] allow communities to express a shared understanding of concepts. Creating ontologies forces communities to systematically address the concepts that they share. For example, if a community uses a concept such as "water pressure," then it may well have a very particular meaning in that community. Defining concepts in ontologies resolves ambiguity and thus forces the community to reach consensus on definitions.

In addition, the process of defining these shared concepts forces communities to be rigorous in their approach to shared concepts. Ontologies require a breadth of structured definition in order to be effective. Definitions or assumptions that are not described in an ontology using structured concepts and explicit terminology are unavailable for inferencing or use. To avoid this, ontology producers must be rigorous and thorough in their production methodology.

Communities can also rely on larger ontological frameworks for use with their domain. By establishing links to "upper ontologies" or "foundation ontologies," which have a broad range of linked concepts already specified, communities can bridge their concepts to larger concepts.

For example, SUMO (Suggested Upper Merged Ontology) is an upper ontology which combines many domain ontologies in order to act as a foundation for the creation of other domain ontologies. It has 20,000 terms and includes ontologies of countries, economy, finance, people, geography, government, and so on. Using these community-supplied definitions decreases the amount of effort needed to create domain ontologies, while increasing their efficacy.

How to cite this Guide

Alexander, P. 2011. "Document and Develop Community Understanding." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/importance/community. Accessed June 23, 2014.

Organize and Share Data

Ontologies [107] can help organize information about a domain, either within a data system or across different data systems. Ontologies achieve this because of their basic features:

- **Ontologies provide a means for defining concepts so that they can be consistently accessed by a computer.**
  - For example, MMI [9] is developing a platforms ontology, which defines different types of oceanographic platforms and their properties (e.g., mobile or not), so that searches for data sets by platform type, or by properties of a platform, can be done automatically across data systems, regardless of the specific terminology used.
- **Ontologies provide hierarchical frameworks for organizing concepts**
  - Classes [225], subclasses, and “is a part of” are examples of the hierarchical organizational framework provided by ontologies for concepts. In the MMI platforms ontology, the superclass concept is "Platform," with its subclass concepts, "AirAndOuterSpaceBasedPlatform," "EarthBasedPlatform," and "WaterBasedPlatform."
- **Ontologies permit the articulation and accessibility of relationships [125] between concepts (instances) and properties.**
  - These relationships can then be used automatically by computers to infer additional associations between concepts. Ontologies can automatically classify concepts according to the properties defined for those concepts. For example, a water-based sensor could be classified as a subclass of a sensor, simply if a property of the water-based sensor is also a property of a sensor. In other words, the property can be used to classify water-based sensors as a type of sensor, without having to manually create water-based sensors as a subclass of sensors.
- **Ontologies can be extended to provide any kind of relationship, or mapping, between individual terms in separate ontologies.**
  - For example, at the MMI workshop, Advancing Domain Vocabularies [311], a sensors working group identified and mapped sensor-related terms from several vocabularies, including those from WHOI [312], MBARI [313], LDEO, SIO [314], TAMU, NGDC, CO-OPS, ACT, and BODC [315]. This work was the precursor to the development of the MMI Sensor Ontology [316] project, the goal of which is to develop a sensor ontology, based on existing vocabularies and the mappings initiated at the MMI workshop.

Through application of these features, ontologies provide a variety of higher level knowledge-management capabilities, each of which helps organize information and knowledge:

- **Ontologies provide consistency to the terms used in metadata records.**
  - Consistency in metadata is essential to keeping information organized, making it discoverable, and enables interoperability [7] between data systems.
- **Ontologies can be used to generate knowledge bases about one or more specific domains(s).**
  - Each ontology represents a set of knowledge about some topic area. By connecting related ontologies, the knowledge framework can be extended to cover a wider domain.
- **Ontologies provide more powerful terms for filling out metadata records, so that they can better represent information.**
  - By formally defining the terms used in metadata records, and enabling those terms to be mapped to other terms relevant to that community, ontologies extend the completeness, precision, computability, and extensibility of the metadata records. (Each term in an ontology can carry with it the context of the entire model, due to more complex semantic statements and the inference capabilities of Description Logics). For example, ontologies can define the values [119] used to complete metadata fields [117] like Keywords. The terms can then be mapped to other vocabularies, and interoperability is facilitated between different metadata records, regardless of the specific terminology used.
- **Ontologies support discovery [113] and understanding through interactive navigation.**
- **The relationships captured by ontologies are analogous to those in topic maps**—they tie together the different terms of the ontology. Visual presentation of these relationships provides a different way to view the knowledge model represented by the ontology, and interactive tools allow it to be easily explored, often with serendipitous results. These explorations can be built into the interfaces used to discover data sets and other scientific materials, allowing searches to be qualified in ways that make sense for the particular subject domain.

How to cite this Guide

Alexander, P. 2011. "Organize and Share Data." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org /guides/vocabs/ont/importance/organizeshare. Accessed June 23, 2014.

Enable Interoperability

Background

There are two major search problems addressed by semantic interoperability [292] between data systems:

1. We cannot find all the data we are seeking.
2. We get too many results and they are difficult to classify.

How Ontologies Can Help

Automated tools can use ontologies for such services as more accurate web search, intelligent software agents, and knowledge management. By formalizing relations between concepts of one or more collections in a machine-readable [112] language, ontologies can facilitate interoperability.

These concept descriptions determine the format in which the information is kept, and establish the actual conceptual information, or semantic content, that is defined in the ontology. Agreements should also be reached about the community and technical processes used to modify the ontology. Finally, ontologies are designed to be computer-usable (also known as "computable") - their format and rules are specified so that the information can be found, exchanged, and applied by computer systems, without additional human intervention.

Some examples of how ontologies can faciltate interoperability:

- Mappings Between Controlled Vocabularies [122]
    - Controlled vocabularies are important, but there is rarely only one controlled vocabulary relevant to a domain of interest. Different funding sources, project purposes, program histories, etc., lead to different controlled vocabularies for a given domain. Mappings between controlled vocabularies, normalized in ontology representation languages such as the Resource Description Framework [317] or the Web Ontology Language (OWL) [318], can consist of identifying terms in each vocabulary as equivalent to, broader than, narrower than, or a subclass of terms in another vocabulary. Such ontology representations and mappings can enhance interoperability between data systems in that the use of specific search terms is no longer necessary. The mappings between terms in different controlled vocabularies used in different data systems can allow the user to find additional information. For example, at the MMI [9] Advancing Domain Vocabularies Workshop in 2005, we demonstrated the enhanced ability to quickly find sea surface temperature data sources (regardless of whether "SST [121]", "sea surface temperature", "Ocean Temperature" variations were used), using an MMI semantic mediation service called Semor [319]. Semor is a semantic mediation service for earth science terminologies. Terminologies are expressed in ontologies following the RDF model [317]. Users can query terminologies using RDF query languages or simple text matching queries. This service helps users discover what a term means and its relationships [125] to other terms.
- Mappings Between Categories/Hierarchies of Concepts

Taxonomies [224] (or other hierarchies) used by different data systems, as well as within a data system, may vary. Ontologies, and mappings between ontologies, can facilitate interoperability between these higher-level categorizations. For example, the Oregon Coastal Atlas and the Marine Irish Digital Atlas, which interoperate as components of an International Coastal Atlas Network [320], use different classifications for grouping their mapping data sets to help users find data sets of interest. MMI is working with this group to create an interoperability prototype [321] between the two atlases, using an upper ontology, as well as mappings between classifications and terms.

How to cite this Guide

Alexander, P. 2011. "Enable Interoperability." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org /guides/vocabs/ont/importance/interoperability. Accessed June 23, 2014.

Using Existing Ontologies

Just as with other types of controlled vocabularies [122], there are many opportunities for communities to use the same ontologies [107]. This can save time and resources and provide a common way to share data among different members of a community.

There are several toolsets that can facilitate your work with ontologies. First, a major hurdle to the adoption of this technology is a simple method by which to discover commonly used ontologies. Search engines and community-run registries and repositories—several of which are listed in the next section—can help.

However, even under the best of circumstances, a pre-existing ontology may need to be adapted for use in particular cases. Fortunately, RDF [187] and OWL [288] provide good mechanisms for extending individual ontologies, allowing them to be modified while still retaining their original meanings and relationships [125].

This section of the guide covers methods for finding ontologies, citing their terms and resources, and extending existing ontologies to work for your needs.

How to cite this Guide

Alexander, P. 2011. "Using Existing Ontologies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org /guides/vocabs/ont/existing. Accessed June 23, 2014.

Finding Ontologies

Ontologies [107] are generally developed around communities, and thus you may already be familiar with some that are being used in your area of interest. However, besides asking within your community, there are several techniques you can use to find ontologies that may be relevant to your needs.

Ontology registries and libraries are excellent ways to discover and share ontologies within a community of users. Semantic-oriented search engines often create indexes of ontologies, even providing ways to work with the ontologies in their index, and general Web search engines like Google index ontologies that can be searched.

General Search

The least sophisticated way of locating ontologies is to use existing Web search technology, like Google, and limit the search results to a particular filetype. In Google, you would do this with the following query:

temperature filetype:owl [288]

This will return results that contain the word temperature, limiting the result set to OWL files.

This method has its limitations, such as Google only recognizes files with extensions. This means that a hosted OWL file without the extension (gcmd [322]-science instead of gcmd-science.owl) will not appear in Google's results.

Registries and Libraries

Ontology registries are generally domain specific, meaning that a community working on similar issues or with similar data will collaborate to provide a service for their community. In their most advanced form, registries can enable interoperability [7] by storing relationships [125] between ontologies, allowing them to act as mediators between different software tools that rely on the relationships between ontologies.

Libraries are generally straight indexes of ontologies, categorized using a set of metadata [5] to provide a catalog of ontologies. However, there are more advanced libraries that allow searching using the semantic query language, SPARQL. Both libraries and registries offer opportunities to find ontologies that may be of use in your work.

The following are registries provided by various ontological communities. Some are applicable to the marine science community, others are provided as examples of how registries function.

MMI Ontology Registry and Repository [323]

MMI has deployed the MMI Ontology Registry and Repository, a version of the BioPortal ontology application, which allows users to access and share ontologies in use in the marine science community. While targeted initially at marine concepts, the MMI Ontology and Registry is open to ontologies from all environmental science fields.

BioPortal [324]

BioPortal is a Web application developed by the National Center for Biomedical Ontologies. It provides the biomedical community a space in which to share and discover ontologies and is the foundation of MMI's Ontology Registry and Repository.

National Cancer Institute Bioportal [325]

The National Cancer Institute (NCI) BioPortal is an older version of the NCBO's BioPortal software. It provides access to the NCI Thesaurus [105], the NCI Metathesaurus, and other select and publicly accessible biomedical terminologies hosted at the NCI.

OntoSelect Ontology Library [326]

The OntoSelect Library provides a catalog of ontologies ranging from cars to plants. There are search and browse functions provided for users and the library tracks metadata for ontologies, including format, language, and domain.

DAML Ontology Library [327]

The DARPA Agent Markup Language Library contains records of ontologies from around the Web. Information from this source may be outdated.

Protégé Ontology Library [328]

The Protégé Library is a flat list of user-provided ontologies hosted on the Protégé project wiki. No metadata is tracked and no search functionality is provided, though the entire list can be searched in a browser.

Search Engines

Unlike Google, which is a general index of the web, there are tools available which allow you to search indexes of ontologies. The power and flexibility of these tools vary, from simple interfaces like Google to more complex products that allow you to run queries in SPARQL.

Swoogle [329]

Swoogle, as the name implies, is an attempt to replicate the usefulness of Google in the semantic web [215] realm. It allows searching for ontologies, documents, and terms, and has an index of over 10,000 ontologies.

ONTOSEARCH2 [330]

ONTOSEARCH2 allows users to search on keyword or use SPARQL for more advanced queries and is under active development. It is currently unknown how many ontologies it includes in its index.

Falcons [331]

Falcons describes itself as, “a keyword-based search engine for the Semantic Web, equipped with browsing capability. Falcons provides keyword-based search for URIs identifying objects, concepts (classes [225] and properties), and documents on the Semantic Web. Falcons also provides a summary for each entity (object, class, property), integrated from all over the Semantic Web.”

Watson [332]

Watson is a keyword search engine for semantic documents that provides great information regarding classes, properties, and individuals in ontologies.

Other Semantic Search Engines

The number of semantic search engines is increasing. You can find a current list, according to the w3c [110] Linking Open Data Task Force [333] and the w3c ESW group [334].

How to cite this Guide

Alexander, P. 2011. "Finding Ontologies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/existing/finding. Accessed June 23, 2014.

Citing Terms

Ontologies [107] and terms within ontologies use addresses so they can be located. To do this, ontologies and their terms are assigned URIs, making them identifiable and useable by people and software.

Where Do URIs Come From?

Unfortunately, there is no one right answer for how URIs are assigned to resources. There are several factors that influence how providers choose to assign URIs to their ontologies:

- Where will the ontology reside? Is it hosted on a Web server, submitted to a registry, or using another method of making it available online? Ontologies submitted to a registry may have a URI assigned by the registry itself, whereas an ontology hosted on a Web server is just like any other file, meaning it resides in a folder hierarchy wherever the provider chose to include it on the server.
- Does the URI include the version identifier for the ontology? As groups make modifications to an ontology, they are versioned to make apparent changes in definitions or structure. Knowing whether or not a version is included in the URI can help you when trying to decide which URI to use for a resource.
- Does the URI help people understand what the resource is? There is some disagreement about this point, and not all ontology providers believe it is important to have URIs that have meaning. However, MMI [9] recommends that semantics be included when assigning URIs as it may provide a way for people to understand a resource simply based on the URI string.

MMI Recommendations

MMI has developed a set of recommendations for ontology providers that are currently in use on the MMI Registry and Repository. The full recommendation [335] is available but is beyond the scope of this document. What follows is a basic description of the URI construction that MMI recommends and uses in its Ontology Registry and Repository.

MMI recommends that URIs are constructed as follows:

URLs for Ontologies

http://{hostDomain}/{ontologiesRoot}/{authority}/{version}/{resourceType}.owl [288] Example: http://mmisw.org/ont/mmi/20081116T071659/sensor [336]

URLs for Concepts (Terms)

http://{hostDomain}/{ontologiesRoot}/{authority}/{version}/{resourceType}/{shortName} Example: http://mmisw.org/ont/mmi/20081116T071659/sensor/ProcessOutput [337]

URIs require no inherent semantics in their construction, meaning that http://aabbcc112233.org [338] is just as valid as http://ontologiesforscience.org [339]. The second representation has much more meaning for a human, though a machine doesn't care either way. MMI recommends that URIs should actually be formulated to have meaning so that a human reading the URI will recognize what it represents and, potentially, what to do with it.

How to cite this Guide

Alexander, P. 2011. "Citing Terms." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/existing/citing. Accessed June 23, 2014.

Extending Existing Ontologies For Your Use

There are many ontologies [75] that have already been created and may be of use in your projects. However, the likelihood of a single, pre-existing ontology that meets all of your requirements may be small. This doesn't mean that you need to start from scratch to create an ontology that satisfies your needs. As an alternative, you can build upon existing ontologies by extending them using standard [157] OWL [288] methods, as demonstrated below.

Namespaces

Namespaces [340] are a feature of XML [341] that were created to provide modularity and allow the re-use of XML code between projects. These same benefits apply to OWL, a type of XML, and namespaces can be declared in your ontology.

When you want to refer to concepts or resources contained within an existing ontology without importing a full version of the external ontology, then you can simply invoke the ontology's namespace, which you can declare in your own ontology. Below is an example of declaring a namespace and then using it in an OWL ontology.

Example Namespace Statement

```
<rdf:rdf xmlns:owl="http://www.w3.org/2002/07/owl#"> </rdf:rdf>
```

This creates the namespace of 'owl', which acts as both a unique identifier and shortcut for the OWL ontology.

Example Namespace Usage

```
<owl:ontology><rdfs:comment>An example ontology that uses OWL</rdfs:comment></owl:ontology>
```

Here, the namespace *owl* that we created above is used with the concept *ontology*, which is defined in the OWL standard. This is the section where the ontology itself is described using metadata [5].

Imports

When creating your own ontology, you may directly import the entirety of a second ontology to be contained within the first. This takes the entire external ontology, including all of its imports and namespaces, and includes them in your ontology in a very literal sense, almost as if you had copied and pasted the text from one file to another.

However, the resources from the external ontology will not actually appear in your OWL file as they do in their original file. All you will see is the import statement, which looks like the following:

```
<owl:imports rdf:resource="http://mmisw.org/ont/cf/20081116T050933/parameter"> </owl:imports>
```

The import statements are read by inferencing engines, RDF stores, and programs like Protege. Once your ontology file is processed via one of the tools, all of the imported ontologies become available for use in queries and inferencing.

How to cite this Guide

Alexander, P. 2011. "Extending Existing Ontologies For Your Use." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/existing/extending. Accessed June 23, 2014.

Creating and Serving Ontologies

If you are thinking about creating your own ontology, you must first do several things to build it, and then a few more tasks to make it useful.

Building An Ontology

The steps to build an ontology depend on what kind of information you want to represent.

Many ontologies [107] are created to capture understanding of a domain. These ontologies relate the different concepts in the domain to each other. A typical resource in such an ontology may say something like:

atmospheric_winds Create sea_surface_waves

Such relationships [125] describe how the domain concepts relate to each other, and make it possible for automated systems to chain those relationships together. The guide Creating a Domain Ontology [342] describes the basic steps to create this type of "domain ontology."

Other ontologies are created to capture terms in a vocabulary or thesaurus [105]. Unlike the more sophisticated relationships in a domain ontology, these term ontologies primarily serve to give each concept a unique reference on the Web, so that other ontologies can specify those concepts in their own relationships. The term ontologies can capture any auxiliary information the user wants to associate with the term, for example a definition (almost always essential in a good term ontology), or a comment. The concepts involved in creating a term ontology are described in Creating Ontologies Using Vocabularies [343].

Promoting an Ontology

Once you have an ontology, you have to decide how you want to make it available to others. The most basic decision is whether you are going to serve your ontology yourself, or let someone else serve it for you. The first is easier, but may be less useful for the community that wants to use your ontology. The tradeoffs are described in Who Serves Your Ontology? [344]

Finally, you will want to consider the best way to make your ontology visible to your community. Similar to "announcing" your Web page to search engines, it can be useful to register your ontology with an ontology search tool, or with a repository. This may have been addressed already if you served your ontology through a repository in the previous step, but additional measures may help make your ontology more accessible and useful. Find out more about these techniques in the guide Registering and Accessing Your Ontology [345].

Summary

As an ontology provider, you are creating very useful information, and providing it to both your immediate user community and to the semantic web [215] community. Following good practices in developing and serving the final product can allow both of these user communities to reap the maximum reward from your efforts, and make your own application of your work easier and more powerful.

How to cite this Guide

Graybeal, J. 2011. "Creating and Serving Ontologies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/provider. Accessed June 23, 2014.

Creating a Domain Ontology

If you want to create a sophisticated ontology describing a domain of interest, there are a few basic steps to get started. Below is a brief outline of the process; the related resources at the end of the guide provide a more thorough treatment.

Community or Solo?

Given your particular goals and knowledge, you will need to decide whether to make the ontology development a community effort. Reasons to do so include greater buy-in (acceptance) and use of the ontology, greater awareness of the work, and access to a larger pool of knowledge.

Reasons you might want to develop the ontology on your own include representing a particular perspective, preserving ownership and credit, and greater efficiency because you eliminate consultation and developer agreement.

Developing a sophisticated domain ontology is usually quite challenging, and at a minimum you will want to have access to a specialist that has some experience developing this kind of information resource. Such an ontology specialist can provide critical guidance and technical support throughout the process.

In the Beginning: Capture Concepts of Interest

As a first step in the process, you'll want to identify key concepts in the domain. Techniques include literature surveys and searches, use-case identification and documentation, analyzing data sets (particular database and file headers, and metadata [5] descriptions), and brainstorming among members of the team.

It is not critical at this stage to get the names or meanings of the terms exactly right. If there is confusion about a particular term, define it or describe the nature of the issue, but don't waste too much time trying to resolve it.

Organize Concepts

For ontology building, it is important to capture not just hierarchies, but meaningful relationships [125] (with verbs) between the different concepts. Using a general tool for organizing concepts (like IHMC CMAP), put terms in a diagram and link them to each other with relationships, creating a large set of linked terms. Many brainstorming and diagramming tools exist to help you organize your concepts.

This is a good time to begin discussing your understanding of the different terms on the page, and to make sure you agree in your understanding of the end document. It isn't necessary yet to have a precise analytical description of all the components—say, at the level of a research paper—but you don’t want to ignore conflicts of views. In some cases, the easiest approach may be to document differences and continue identifying relationships.

Formalize Relationships, Classes, Properties and Instances, and Subclass Relations

Certain relations will be apparent in almost any concept diagram from the previous step: one thing "is a" thing of some other type; this thing "has" those things. Key to an effective ontology will be identifying which terms are really properties of another (a rock has a shear strength, or a fluid has a boiling point), and which are general concepts (cars) as opposed to instances (Honda Civics).

At the same time, you should identify key relationships that are necessary to create your ontology, and define them in terms other semantic tools can use (is it a transitive relationship? is it symmetric?).

Another type of relationship is that of a class to its subclass. While a Honda Civic is certainly an instance of "car types," the Honda Civic in my driveway is an instance of the sub-class "Honda Civic model." So, whether a concept is a class, a lower-level subclass, or an instance, is not always straightforward.

This can be the trickiest part of building an ontology, as the classification of concepts into the different categories can depend not just on subtle judgments of the usefulness of each classification, but also on the purpose to which the ontology will be put. For example, classifications appropriate for a samples database may not be very effective for a training module. An expert knowledge engineer can contribute to this stage of discussions.

Capturing the Information in a Knowledge Model

Depending on the situation, discussions in the last section can take place on a white board, in a drawing or concept mapping tool, or in an ontology-building application such as the ontology editors Protégé [346] or TopBraid [347]. The final step in the initial process, at least until iteration begins, is to capture the discussions as thoroughly and accurately as possible using an ontology editor.

This process will either strengthen, or question the knowledge model realized in the ontology. These discussions can be represented as additional relationships, which add inferences into the model. The added relationships and inferences will either support the consistency and usefulness of the ontology, or identify problems that need to be resolved.

Iterations

As new information is added to the existing model—or the existing model and its inferences are reviewed and used by other systems—discrepancies and issues inevitably arise. A process is necessary by which the model owner (individual or community) can review the work and update it. This can be expected to continue indefinitely for more complex models, and even more so for those that represent cutting-edge research. Getting the knowledge in an ontology "just right" is usually not a goal for the short term, but with increasing maturity and feedback the ontology can become increasingly consistent, powerful, and reusable.

References

The following references provide much more information about creating ontologies [107].

- Ontology Tutorial [348] (Natasha Noy, Deborah McGuinness)
- Guided Tour of Ontology [349] (John F Sowa)
- Ontology [350] (John F Sowa)
- Introduction to Ontologies and Semantic Web [351], (Mark Obitko)

How to cite this Guide

Graybeal, J. 2011. "Creating a Domain Ontology." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/provider/createdomainont. Accessed June 23, 2014.

Creating Ontologies Using Vocabularies

For flat or hierarchical vocabularies [130], it is often straightforward to create a corresponding ontology. The development of an ontology will involve ensuring the following: format, classes [225]/subclasses, instances, and relationships [125]. Here are a few key things to understand before you begin creating the ontology.

Deciding on Term Identifiers: Opaque or Meaningful

In general, terms are represented in a vocabulary using a string of alphanumeric characters. These may be meaningful and describe the term (e.g., temperature), or they may be opaque and represent the term (e.g., 729402c).

A first approach to creating unique references to terms on the Web might be to give them an identifier that contains their name, for example http://mmisw.org/mmi/examples/sea_surface_temperature [352]. Here 'sea_surface_temperature' is the name associated with a given concept, and we can define what it means, and life is good, right?

Unfortunately, it is in the nature of language, and of scientific terminology, that meanings of terms change over time. This is true even in such terminology-focused domains as species classification. Species names change, species identifications change, even the way species are classified into higher groups is subject to major change. So, we need to appreciate that the thing we called 'sea surface temperature' 50 years ago may have a different name next year, like 'sea surface foundation temperature.'

Most ontologists have determined that the way to avoid this problem is to create a unique identifier—specifically for the Web, a Uniform Resource Identifier, or URI—for the concept that is of interest, and to make that identifier 'opaque,' that is, without any meaningful concepts embedded in it. The identifier may be a number, or a code, or a random string, but it has an associated preferred label [180], and a definition, which clarifies the concept of the URI. If the way people refer to the item changes, say from 'frisbee' to 'flying disc,' then the label can easily be changed, while the concept's URI stays the same.

Ontology providers can use either method. Both have strengths and weaknesses. The URI method provides a powerful abstraction, but it requires ontology users to know the representative string of characters. The meaningful term method is instantly clear to users, but it must accommodate changes in terminology and meaning.

For the creator of a vocabulary, determining an approach may not require deep understanding of the two options. It is easy enough to use numbers or codes for your vocabulary entities, and in fact many communities that negotiate shared vocabularies find that using numbers is the only way to avoid unending arguments about which term to use. On the other hand, if you are creating a local vocabulary that is unlikely to have extensive persistence or community use, the simplicity of meaningful identifiers can be a powerful incentive toward standardization.

There are two other reasons to consider opaque, or at least non-literal, terms for your vocabulary. The first is the ability to specify the label as accurately as you want. If you use the term as part of your identifier, it becomes very awkward to create a usable URL [279] out of a term like 'Crutchfield Jacob's Syndrome.' Terms that include spaces, hyphens, pound signs, slashes, accent marks, or other non-ASCII [232] and non-Roman characters work poorly as part of a URL in a browser.

The second simple reason to avoid using terms as identifiers is when your terms have multiple meanings. Since identifiers must be unique, it may make a lot more sense to use numbers than to use successively longer discriminators, as in 'sea_surface_temperature_when_moving_measured_by_thermistor_uncalibrated.' You get the idea.

This guide suggests that you choose the approach you consider most appropriate to your community, considering its size, longevity, diversity, and visibility to the greater science community, and your ability to support codes in your data systems. If you decide to use terms as meaningful identifiers, be aware that eventually there may well be repercussions associated with that approach.

Other Information to Include

We know it is important to include the string used as the identifier, a preferred label (if your term is not the preferred label), and a definition. What other information should be included with your vocabulary and its terms?

At the vocabulary level, a number of metadata [5] items are worth capturing: who created it, the date, its purpose, its principal topic, and maybe the terms under which others can use it. There are metadata standards like Dublin Core and the Ontology Metadata Vocabulary, as well as extensions like MMI's Metadata Vocabulary that can help you decide what to include.

At the term level, what other information is worth including? Information about the semantic content of the term may be useful. For example, a complete spelling if it is an acronym, or a URL for the home page if it is an organization. Alternate labels may be of value [119], as well as relationships to terms in another vocabulary. These are not recommendations, simply suggestions to consider.

The tools that work with controlled vocabularies [122] and ontologies [107] are very term oriented, so you should not include concepts unrelated to the meaning of the term itself. The format and units of a data item may seem useful, but prevent you from using that term and definition in other contexts where the format and units might be different, or the word might not represent a data item at all. To think of it another way, the vocabulary is not intended to define all the metadata for a data variable, simply the language used to name the data variable.

It is simple to add information to your vocabulary, but if the list of information is more than a few items, or you want to express the relationship of your terms to each other, you may be better off working directly with ontology tools, as described in previous pages of the MMI Guides.

Translating to an Ontology

A good starting point for your vocabulary is to have each type of information about the terms—identifying name, definition, and other pieces of information—stored in a separate column, separated by tabs or commas. A spreadsheet can easily generate data in either of these formats. Remember that the identifying name has to be unique, without spaces or unusual characters (keeping it to A-Z, a-z, 0-9, and _ is a good practice). Given this, how do we make an ontology?

A basic ontology has a simple format, so a simple pattern substitution could turn your vocabulary list into a credible term ontology. Examine a few other term ontologies to identify many of the basic patterns, if you want to experiment when creating your own. An ontology editing tool like Protégé [353] (free) can help you confirm the ontology is the way you want it, or can help you create the ontology from scratch.

With the voc2rdf [354] tool introduced by MMI, there is another way to move from a vocabulary list to an ontology. With this service, it is possible to enter the comma- or tab-delimited text into a dialog box, and press a button to have it converted to an ontology. If you want to get off to a quick start, this is a reasonable way to proceed, and you have the option of working with the resulting ontology, or committing it directly to a repository.

If your vocabulary is at all complex, with hierarchies or internal relationships, you probably should start with an ontology tool, and some of the basic instructions on ontologies here and elsewhere. But for simple vocabularies, some simple approaches may work well.

How to cite this Guide

Graybeal, J. 2011. "Creating Ontologies Using Vocabularies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/provider/createvocabont. Accessed June 23, 2014.

Who Serves Your Ontology?

Should you serve your ontology yourself, or via a repository?

When you have an ontology, you have to choose whether to present it yourself, or via one or more third-party repositories. There are good reasons to do it either way, as described below. Additional considerations are described in the section on Registering and Accessing Ontologies [82].

A significant consideration is that ontologies that are published on the Web can be harvested by various registries, thereby providing some of the advanced services that working through a registry can provide. This is discussed further in the Summary below.

Serving An Ontology Yourself

Advantages to self-publishing include simplicity, control, speed, and namespace identity. Any or all of these gains may be sufficiently important to overcome the advantages offered by an organized publication service.

Publishing an ontology can be as simple as making a document called {someName}.owl [288] available on the Web. This solution maximizes simplicity, control, and speed for the ontology provider, since making a file available on the Web is trivial for even casual Web publishers. The entire publication process is within the control of the provider.

A second strategic value [119] of self-publishing is that the Uniform Resource Identifiers (URIs) associated with the ontology resource, and its terms, can be local to the provider and resolved at the provider's website. So, for example, the provider can declare that the namespace of the ontology—the string used as a prefix for the terms in the ontology—can be in the provider's domain, such as http://mywebsite.com/ontologies/ [355].

With this namespace arrangement, all the term identifiers will be visible from the provider's domain, and the provider can decide how to respond if someone enters one of those terms into a Web browser or ontology tool. This provides a complete level of control over the entire user experience with the ontology.

Serving An Ontology Through a Third Party

A complementary set of advantages are obtained by letting an ontology provider serve your ontology. Both organizational and technical strength is offered by most of these ontology hosts, since they have to serve multiple ontologies. At the same time, the field [117] is so new that each provider often provides a relatively unique set of services, so the value may be diffuse depending on exactly what services you are looking for.

The most visible value offered by an ontology repository is the services (functionality) built in to the repository. These services run from basic to advanced:

- storage and backup
- registration and Web visibility, with association with other similar ontologies
- tools to ensure the ontology is viable and follows good practices
- tools and services to analyze, search, cross-reference, and track usage of your ontologies
- metadata [5] guidance, entry, and maintenance
- human and online advice on ontology creation
- automatic URI generation
- automatic URI resolution (dereferencing)

The last two services above require more discussion. An ontology's terms should be at least named on the Web, by creating labels for each term resource. These labels can be a Uniform Resource Identifier (URI) in any form—a Web address, a Uniform Resource Name (URN [356]), a Universal Unique Identification (UUID), or a number of other forms. The author or publisher of the ontology should declare the URI that is associated with each term to avoid ambiguity about the identifier. (Multiple identifiers for a single semantic resource is generally A Bad Thing.)

Another nice feature is to provide a Uniform Resource Locator, or URL [279], a Web address at which more information about the term can be obtained. While it is not necessary that the URL and the resource URI be the same—some experts even discourage this practice—there are significant social advantages to using the same string for both purposes. If you want to achieve this social purpose, it means implementing some additional capabilities to name and resolve the ontology terms you have created.

An ontology service can offer social value as well. By providing organizational stability and permanence, the service provides preservation for your ontology over time, which may not be possible with your local Web server and organization. And by archiving old versions of the ontology and making them available and comparable, and ontology repository preserves the history [108]

The record of how a particular value or record came to be. Provenance can include things like when, by whom, and how the item was created or modified.

of your work.

Registering Through a Third Party

A third approach is registering your ontology with an ontology registry, while continuing to serve the document yourself. This is similar to registering your Web pages with search engines. It can be a suitable middle ground to achieve your goals, if the services provided by the ontology registry complement the services you are capable of providing. Further discussion of this is available in the next section.

Summary

If you are doing very basic ontologies for your own use, or the use of a narrow community, publishing it yourself with minimal services is a reasonable tradeoff. If you have obtained a basic review of the ontology you are publishing, it should be possible to make a useful asset publicly available with fairly limited resources.

On the other hand, if your ontology is at all likely to be a community resource, you need to be aware of the value that can be added by using community services. By considering which capabilities are important to your needs (and, ideally, which capabilities will provide the most community advantage), you can often increase the value of your ontology both for yourself and for a large collection of other users. These improvements can often be achieved with very little additional effort and can save significant effort later.

How to cite this Guide

Graybeal, J. 2011. "Who Serves Your Ontology?." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/provider/servingontologies. Accessed June 23, 2014.

Registering and Accessing Your Ontology

Once you have your ontology in hand, you need to tell the world about it. You also need a way for the world to easily access your information.

Registering an Ontology

Broadly, there are four modes in which you can register an ontology: submitting it to a central repository, where it is stored and served; registering it in a central repository, which may thoroughly index or even cache it; registering it in an ontology search engine, which will index certain metadata [5] about it, but will not store the original ontology for you; and registering it in regular Web search engines, which index it and make it available without caring that it is an ontology. Each has advantages.

Submitting to a central repository

If you submit your ontology to an effective central repository, this may avoid other steps. The repository should make your ontology visible on the Web as an OWL [357] file, and may even register it with other search engines. The submission also enables the repository to perform other operations like versioning. Of course, this is only effective if all your versions are submitted!

Registering in a central ontology registry

You may want to just register your ontology with a central registry so that the registry can index it and knows about it, while not actually serving your ontology. This is useful if you want the URIs for the terms in your ontology to be in your own namespace, and you want to provide dereferencing (Web responses) for those URIs yourself.

Some registries may perform more advanced indexing and computations with your ontology, making them closer in function to a central repository. However, they will still not be able to provide resolution (dereferencing) services for term URLs that are in your domain.

While some registries may revisit the ontology every so often to see if it has changed, you should confirm this, and consider re-registering the ontology each time it changes.

Registering in an ontology search engine

There may be little distinction between some ontology search engines and ontology registries because these capabilities are still evolving on the Web. As used here, the difference is that a central registry provides more advanced services than simply the ability to find information from the ontology.

An ontology search engine will index your ontology, understanding its ontological components, much like a registry. It might also cache your ontology, but not as a primary service. Compare this to the Web search engine, described below.

The best reason to register with an ontology search engine is that it is dedicated to your particular domain, e.g., environmental science or volcanology.

Ontology search engines may be aware of some ontologies [107] that are not available via the Web and may provide additional categorization functions that are specific to ontologies when compared to regular search engines.

Registering with a Web search engine

Most Web search engines support registration of files or websites. One advantage of registering your ontology with Web search engines is that it becomes visible to the widest possible community. At the same time, anyone who wants to search for terms in ontologies can limit their searches to files ending in .owl, in those search engines that support this kind of advanced search.

If you publish multiple ontologies, some of which may not be linked by other websites, it will be especially important to directly register each ontology with Web search engines.

Enabling Access to your Ontology

Once your ontology is known to the Web, how can people and machines access it? This involves two different sets of customers: those using Web browsers, and those using semantic tools. The former wants to see nicely formatted HTML [162], while the latter wants to see XML [131] in the form of RDF [187]. Obviously you will have to have a Web server of some sort in place, providing the HTML and the RDF responses.

Some basic information on serving your ontology, for example useful naming conventions, is offered in the Guide for Ontology Providers [358]. The page on Constructing URIs for Ontologies [359] is expecially relevant. More advanced information about serving RDF, even at the same URL as the HTML, is provided by the World Wide Web Consortium (W3C [110]) in its Best Practice Recipes for Publishing RDF Vocabularies [360]. This document contains advice for a wide range of web server configurations. Both of these documents are recommended reading if you are going to have to deal with ontology publication on a regular basis. And if this is just an occasional activity, by all means consider using an ontology repository to take care of the details for you.

How to cite this Guide

Graybeal, J. 2011. "Registering and Accessing Your Ontology." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/provider/registeraccessont. Accessed June 23, 2014.

Core Technologies for Ontologies

The use of ontologies [107] for data interoperability [7] and Semantic Web [215] technologies relies on a core set of standards developed by the World Wide Web Consortium (w3c) [361]. Using these standards, technology implementers everywhere can deploy solutions that work not only in their individual instance [160] but can work with software and data created and provided by others.

The core technologies that enable this are Extensible Markup Language (XML [131]), Resource Description Framework (RDF [187]), and the Web Ontology Language (OWL [288]). Each of these plays a role in the creation, storage, and use of ontologies.

XML is the backbone of both RDF and OWL, as each of them are built on the extensible programming language platform that is the essence of XML. RDF provides a specialized extension [286] of XML that allows for the description of resources in ontologies using a standardized set of syntax.

OWL builds upon RDF by adding an ontological layer, meaning that the resources described by RDF are capable of being classified and defined by OWL.

In this section of the guide, we will introduce you to these core technologies and provide a brief description of each. In addition, we discuss how to store and access ontologies using several methods. We provide an overview of common software libraries that are used to work with ontologies and, finally, we discuss how computers can analyze ontologies to draw inferences using a specialized technology called inference or reasoning engines.

How to cite this Guide

Alexander, P. 2011. "Core Technologies for Ontologies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/coretech. Accessed June 23, 2014.

Ontology Standards

Ontologies [107] written using the Web Ontology Language are built on a set of standards that are developed by an international consortium known as the World Wide Web Consortium (W3C) [362]. The W3C is made up of member organizations, paid staff, and interested members of the public, and has produced a large number of standards on Internet protocols [132]. Each of the standards is open and publicly available. Below, we briefly discuss the relevant standards that are being developed, cover their current status, and explain their relationship [125] to ontologies.

Extensible Markup Language (XML)

XML is a specification that allows people to create custom markup languages for specific needs. Roughly, it can be considered a programming language for creating other programming languages.

Another markup language, HTML [162], is used throughout the Internet and is loosely based on markup languages used in the publishing industry that indicate where text should be made bold, italic, or formatted when being set for printing. Similarly, XML provides a standardized way for anyone to create a specialized language for individual needs. An example of such a language in the marine sciences is SensorML [363], a language for "describing sensor and measurement processes."

When someone wants to create a language using XML, they need to define its schema. The schema works in conjunction with existing XML syntax and rules and helps to define the language being created. For example, SensorML's base schema [364] includes things such as definitions for elements and groups, similar to individuals and classes [225] in ontologies, which can be used by applications that implement the SensorML language to store, retrieve, and analyze data. Because SensorML conforms to the syntax and rules of XML, any application that knows how to read XML data will be able to work with the language. However, unless an application was written to specifically work with SensorML it won't be able to take advantage of the unique elements and groups that the language defines. This is because XML lacks semantics—something that is corrected with the implementation of the Resource Description Framework and Web Ontology Language.

Related Links

- W3C Specifications for XML [365]
- XML Tutorial from W3Schools [366]

Resource Description Framework (RDF)

RDF is a data specification used to make statements about resources using subject-predicate-object statements called *triples*. Each of the subject, predicate, and object terms is typically a Web resource (though the object can also be a constant). In a triple, the predicate expresses the relationship between the subject and object.

To provide a plain language example, let's examine the statement: “The car has the color red.” “The car” is the subject, “has the color” is the predicate, and “red” is the object. Triples are a powerful model for describing resources.

RDF is a stable standard as the tasks of the W3C RDF Core Working Group were completed in 2004.

Related Links

- W3C Specifications for RDF [367]
- PlanetRDF Guide [368]

Web Ontology Language (OWL)

OWL is a suite of knowledge representation languages that are used to construct ontologies. These languages include (from simpler to more complex): OWL Lite (rarely used), OWL DL [369], and OWL Full. All are based on the RDF/XML formats, described above. Though the languages are similar, the more advanced forms have features unavailable to the simpler ones, preventing full interoperability [7] between the languages.

OWL ontologies can contain classes, which are used to categorize concepts with similar characteristics. These classes can be defined and restricted by *axioms*, statements that are considered true and which lay the groundwork for inferencing. OWL ontologies may also include *individuals*, sometimes referred to as instances, for example, a list of terms in a vocabulary [130]. These instances are grouped together as class extensions, which are related to particular classes without defining them. Therefore, two or more classes could share a class extension, meaning that they would be related to the same group of instances.

Basic inferencing can be performed based on the characteristics of defined relations. For example, assume the relationship "is larger than" is defined as the inverse of the relationship "is smaller than." Given a statement such as, "A car is larger than a bicycle," computer software can use the inverse relationship to determine that a bicycle is smaller than a car. Transitivity and symmetry are other primary ontological characteristics that can be defined for relations.

More advanced inferencing is possible. For example, let's consider an ontology that describes human beings. If the class we use to group humans in this ontology is described with an axiom that indicates all humans have the property "hasParent" in combination with the property "hasMother," a computer could infer that every human being must have a parent and a mother.

OWL 1 has been finalized and is currently in use. In October 2007, a W3C working group was formed to extend OWL. The new version, OWL 2, is already in use in some applications.

Related Links

- W3C OWL Reference [370]
- W3C OWL Guide [371]

How to cite this Guide

Alexander, P. 2011. "Ontology Standards." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org /guides/vocabs/ont/coretech/standards. Accessed June 23, 2014.

Storage and Access Fundamentals

The most basic way to store and access ontological information is by using regular XML [131] files written using RDF [187] and OWL [288] syntax. These files can be stored on a computer that you access locally or they can be made accessible to others by hosting them on a public Web server. OWL files are commonly used and can be opened and saved by editing programs such as Protege and TopBraid.

Ontologies [107] are often stored online, either on a Web server, repository, or possibly by programatic means. These ontologies become available via a URL [279] and, when used with repositories, can include advanced features like versioning and URI [212] generation. MMI [9] provides more information about registering and accessing ontologies for those interested in an ontology providers perspective [78].

Triple Stores

Another common method for storing ontologies involves the use of a database-like technology called a triple store. RDF documents can be expressed in these stores in their triple state, meaning that the information is stored as statements including subject, predicate, and object. There is a lot of debate about just how effective triple-stores are, especially in relation to traditional database technologies, and it remains to be seen exactly how they will emerge as the technology matures.

Triple stores are optimized for handling the type of RDF data that make up ontologies and can be queried to discover information about data and its relationships [125]. SPARQL is a query language, similar to SQL, that can be used to write queries specific to RDF data, meaning that you can query using semantic data provided in the ontologies you are working with.

SPARQL

SPARQL is an emerging standard [157] method for making queries about RDF triples against RDF data stores. Unlike SQL, SPARQL is RDF-aware, meaning that it can take advantage of the definitions and descriptions provided in ontologies.

For example, you could limit your search to "all things that are the color red" or "all documents created by President Lincoln and anyone who worked with President Lincoln." Of course, the ontologies and datasets you are querying will need to be properly described and defined for this to work.

SPARQL has the capabilities to support many more complex queries than the simple examples above. SPARQL contains specific provisions for use in obtaining descriptions about particular RDF resources. The query language also supports advanced methods for filtering results, including the use of regular expressions.

How to cite this Guide

Alexander, P. 2011. "Storage and Access Fundamentals." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/coretech/storage. Accessed June 23, 2014.

Software Libraries

There are many commercial and open source software libraries and APIs for working with XML [131], RDF [187], and OWL [288] technologies. This allows you to easily incorporate their use in your applications or develop tools that leverage semantic technology [305] in your particular domain. Below, we briefly describe three of the many software libraries available and present resources for discovering others. All information is current as of last publication (see citation).

Protégé-OWL API [372]

- Open Source
- Java-based
- Active development

The Protégé-OWL API was developed by the Protégé community and development team at Stanford. Using the API, it is possible to easily develop Java-based software that can open, save, query, and reason against ontology files. The Protégé user community is also quite large, with their main mailing list membership totaling close to 200,000.

Jena Java Framework [373]

- Open Source
- Java-based
- Active development

Jena is a very popular library for working with ontology data in Java. Like Protégé, it allows developers to easily include ontology capabilities in their software, including reasoning, SPARQL querying, APIs for RDF and OWL, and a storage mechanism for holding RDF information.

RDFReactor [374]

- Open Source
- Java-based
- Active development

Instead of providing software libraries or APIs for working with RDF or OWL information, RDFReactor takes the approach of converting these formats into something that software developers are already familiar with: objects. RDFReactor is a Java-based library that allows you to work with RDF data as if it were traditional object-oriented data. This familiarity may lessen the learning curve for those getting familiar with ontologies [107].

Talis Platform [375]

- Open Source
- Web service [221]
- Active development

The Talis Platform takes a different approach to providing tools that enable developers to easily work with ontology data. Instead of providing a downloadable API for use with a particular software language, they have created a web-based service that provides mechanisms to store, index, search, and augment ontology data.

Other Libraries, APIs and Platforms [376]

There are many libraries, frameworks, and APIs for developers to build upon when working with the semantic web [215] and many more are being developed. This guide only covers a small portion of the available solutions and we encourage you to leave feedback regarding software libraries that you find useful. If you would like to explore other options, the W3C [110] has an expansive collection [376] available for review.

How to cite this Guide

Alexander, P. 2011. "Software Libraries." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/coretech/softwarelibs. Accessed June 23, 2014.

Inference Engines

Inference engines, also known as semantic reasoners or reasoning engines, provide the technological glue that allows computer mediation to take place. Inferencing is the process where computers draw connections between pieces of data or metadata [5] that have been previously described by a human. In the semantic web [215], these data points generally reside in an ontology. Inferencing is used in other areas of computer science as well, though generally in more limited and less distributed data sets.

Inference engines rely on a formalized language of logic to implement rules about the world in which inferencing takes place. This language is distinct from ontologies [107] and uses *concepts*, *roles*, and *individuals* to describe the relationship [125] between objects in the world. This allows computer software to find implicit facts using the explicit descriptions in your ontologies.

Potentially, this could uncover information that you aren't aware of or help to confirm or make clear assumptions you already have about your particular domain.

There are many inferencing engines, each with its particular specialty or purpose. Many of the available software libraries [377] also contain some functionality for handling reasoning and inferencing. The W3C [110] maintains a list of popular inferencing engines [378] along with brief descriptions.

How to cite this Guide

Alexander, P. 2011. "Inference Engines." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/coretech/inferencing. Accessed June 23, 2014.

Ontology Tutorials

Tutorials developed by MMI for working with ontologies and ontological tools

- MMI Ontology Creation Guidance [379]
- Using Protege to Edit OWL Ontologies [380]
- Ontology Web Services Tutorial [381]
- URI and Namespace Tutorial [382]

How to cite this Guide

Galbraith, N. 2011. "Ontology Tutorials." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/vocabs/ont/onttut. Accessed June 23, 2014.

Technical Tools

Technical Tools

Creating metadata [5] can be a challenging task. Fortunately there are a variety of tools to assist with such tasks as creating metadata, editing metadata, and crosswalking between different standards and ontologies [107].

A set of guides about metadata tools has been identified as desired by MMI [9] readers. MMI is a community effort. If you would like to help with organizing and writing guides about metadata-related tools, please let us know [278].

For more information on this potential new initiative, please see the MMI DIVE webpage [383].

How to cite this Guide

Neiswender, C. 2011. "Technical Tools." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides /technicaltools. Accessed June 23, 2014.

What is a Tool?

Coming soon... Watch for updates! Tabela [384]

How to cite this Guide

Neiswender, C. 2011. "What is a Tool?." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/technicaltools/whatistool_spammed. Accessed June 23, 2014.

The Importance of Tools

Coming soon... Watch for updates!

How to cite this Guide

Neiswender, C. 2009. "The Importance of Tools." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/technicaltools/importancetools. Accessed June 23, 2014.

Choosing a Tool

Coming soon... Watch for updates!

How to cite this Guide

Neiswender, C. 2009. "Choosing a Tool." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/technicaltools/choosingtool. Accessed June 23, 2014.

Functions

Coming soon... Watch for updates!

How to cite this Guide

Neiswender, C. 2009. "Functions." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/technicaltools/choosingtool/toolfunctions. Accessed June 23, 2014.

Licensing

Coming soon... Watch for updates!

How to cite this Guide

Neiswender, C. 2009. "Licensing." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/technicaltools/choosingtool/toollicensing. Accessed June 23, 2014.

Implementations

Coming soon... Watch for updates!

How to cite this Guide

Neiswender, C. 2009. "Implementations." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/technicaltools/choosingtool/toolimplement. Accessed June 23, 2014.

Case Studies

Practical examples of metadata system implementation and use

DIGARCH Cruise Harvest [385]

Scripps Institution of Oceanography (SIO [314]) / Woods Hole Oceanographic Institution (WHOI [312]) project to archive cruise data in an interoperable [7] environment, using tools developed at SIO

Controlled Vocabularies (CV) for Metadata Harvesting [386]

Scripps Institution of Oceanography (SIO) experience using database technology to harvest and correct erroneous metadata

Controlled Vocabulary (CV) Dictionary [387]

Draft of the Controlled Vocabulary (CV) Dictionary Case Study. Prepared by Stephen P. Miller, 2007-06-05

Attachment	Size
Controlled Vocabularies (CV) for Metadata Harvesting (PDF) [388]	77.19 KB
Controlled Vocabulary (CV) Dictionary (PDF) [389]	96.46 KB
DIGARCH Cruise Harvest (PDF) [390]	79.36 KB

How to cite this Guide

Neiswender, C., Miller, S.P. 2009. "Case Studies." In The MMI Guides: Navigating the World of Marine Metadata. http://marinemetadata.org/guides/casestudies. Accessed June 23, 2014.

Additional Resources

Guides, tutorials and informative resources useful for the marine data community.

If you know of a resource not included on this list, please let us know [2], or submit the reference online [391]!

Reference	Description	Guide Reference Topics
Advancing ecological research with Ontologies (a review paper) [392]	A review paper on the use of ontologies in ecological research	Convention Topics [393] Resource Discovery [394] Species [395] Usage Example Topics [396]
GEON Summer Institute [397]	Metadata-related materials from the Cyberinfrastructure Summer Institute, put on by GEON each year. 2005 includes presentations on Semantic Web and Knowledge Representation, Ontologies in GEON, and many other (non-metadata) data systems related topics, like web services.	Guide Reference Topics [398] Metadata [399] Controlled Vocabularies [400] Tools [401]
Getty Library Introduction to Metadata [402]	An online publication devoted to metadata, its types and uses, and how it can improve access to digital resources.	Guide Reference Topics [398] Metadata [399]
GSDI Data Management Cookbook [403]	Data Management Cookbook for geospatial data. Old but well constructed example.	Guide Reference Topics [398] Metadata [399]
ICES/IOC SGXML Final Report [404]	The final report (including recommendations relevant to metadata practice in the marine domain)of the International Council for the Exploration of the Seas / Intergovernnmental Oceanographic (ICES/IOC) Study Group on the Development of Marine Data Exchange Systems using XML (SGXML).	Guide Reference Topics [398] Metadata [399] Tools [401]
NISO Primer: Understanding Metadata [126]	A primer introduction to metadata from NISO (July 24, 2004)	Guide Reference Topics [398] Metadata [399] Metadata Standards, Extensions and Profiles [405]
NOAA's CSC Metadata Workshop Materials [406]	This online guide, available from NOAA's Coastal Services Center, includes information on metadata standards, getting started, metadata tools, metadata training, and metadata links.	Guide Reference Topics [398] Metadata [399] Metadata Standards, Extensions and Profiles [405] Tools [401]
NSDL Metadata Primer [407]	Metadata primer guide written for the National Science Digital Library (NSDL) by Diane Hillmann's.	Guide Reference Topics [398] Metadata [399] Metadata Standards, Extensions and Profiles [405]
NSGIC Metadata Primer [408]	A "How To" Guide on Metadata Implementation for digital spatial data	Guidance [409] Guide Reference Topics [398] Metadata [399] Initiative Topics [410]
OCGC's Metadata Demystified [411]	What are metadata and why are they important?	Guide Reference Topics [398] Metadata [399] Metadata Standards, Extensions and Profiles [405]
Scholarly Citation of Data [412]	A recent paper proposing a standard for the scholarly citation of quantitative data.	Guide Reference Topics [398] Metadata [399] Metadata Standards, Extensions and Profiles [405]
Stanford Ontology Creation Guide [413]	A guide about ontology development published by the creators of Protege at Stanford University.	Guide Reference Topics [398] Controlled Vocabularies [400]
UNESCO IOC Ocean Teacher [414]	Provides basic technical training on establishing oceanographic data systems.	Guide Reference Topics [398] Metadata [399]
Virtual Data Center Citation [415]	A (somewhat) specific implementation of the Altman-King recommendation for data citations.	Guide Reference Topics [398] Metadata [399] Metadata Standards, Extensions and Profiles [405]
W3C Best Practices for Semantic Web (2006) [416]	A list of best practices developed for the semantic web	Usage Example Topics [396]
W3C: URIs, URLs, and URNs [417]	W3C description of the relationship between URIs, URLs, and URNs	Guide Reference Topics [398] Tools [401]

Glossary of Metadata Terminology

A glossary of the terms used on the Marine Metadata Interoperability project web site.

This glossary provides metadata-related definitions for the terms used on this web site.

In some cases definitions used by the larger marine metadata community may be different than those used by this site; in such cases, the difference is indicated. In many cases this page provides a metadata-centric definition for a term that has wider meetings (e.g., 'dictionary'); such cases begin with the phrase 'In the context of metadata'.

Metadata-Related Definitions

Add term [418] Edit guides glossary [419] Glossary settings [420]

A

Authority File edit term [437]
: A type of flat controlled vocabulary that consists of a list of labels and terms which can be used for establishing the acceptable content, for example a metadata element or database field. Related Guide [43] ***Synonyms**: Authority Files*

C

Clearinghouse edit term [438]
: A service available via the internet that provides a catalog of resources. A clearinghouse usually emphasizes discovery of resources, particularly data sets.

Code List edit term [439]
: A type of flat controlled vocabulary consisting of a set of codes and their meanings, in use in a specific project. Related Guide [46] ***Synonyms**: Code Lists*

Content Standard edit term [440]
: A list or hierarchy of required metadata elements to be included in the metadata description. Related Guide [177] ***Synonyms**: Content Standards*

Controlled Vocabulary edit term [441]
: A managed list of terms. In the context of vocabularies, management typically includes careful selection of terms, maintenance of terms over time (i.e. addition, deprecation, modification), and presentation of the vocabulary in an accessible format. Related Guide [89] ***Synonyms**: Controlled Vocabularies, CV, CVs*

Crosswalk edit term [442]
: Documents that map metadata elements between different metadata standards. Related Guide [32] ***Synonyms**: Metadata Crosswalk, Vocabulary Mappings, Vocabulary Mapping, Crosswalk*

Crosswalking Rule edit term [443]
: In the context of crosswalking, rules are a process which define how to deal consistently with complex element mappings. Rules are created and applied during the mapping of elements from the source schema to the target schema, when one-to-one relationships between schema elements do not exist. Related Guide [35]

 ***Synonyms**: Crosswalking Rules*

D

Dictionary edit term [444]
: In the context of metadata, a dictionary is a type of controlled flat vocabulary, which provides a list of metadata terms, definitions and additional information within a specific domain. Related Guide [44] ***Synonyms**: Dictionaries*

Discovery edit term [445]
: Use of metadata values or vocabularies to find metadata or data sets. Related Guide [189] ***Synonyms**: Discovery Metadata, Discovery Vocabulary, Discovery Vocabularies*

F

Flat Vocabulary edit term [446]
: A managed list of acceptable metadata terms that associates acceptable values with particular metadata elements. Flat vocabularies include authority files, glossaries, dictionaries, code lists, and gazetteers. Related Guide [235] ***Synonyms**: Flat Vocabularies*

Formal Metadata edit term [447]
: Metadata that conforms to a specific standard, with consistent collection criteria, terminology and structure. Related Guide [26]

Format Standard edit term [448]
: A description of the digital storage and structural requirements of metadata which assures that different software programs are able to read or query the data. Related Guide [179] ***Synonyms**: Format Standards*

G

Gazetteer edit term [449]
: In the context of metadata, a gazetteer is a very specific type of flat controlled vocabulary - a geographic term list. Related Guide [45] ***Synonyms**: Gazetteers*

Glossary edit term [450]
: A type of flat controlled vocabulary containing a list of terms in a particular domain of knowledge with the definitions for those terms. Related Guide [14] ***Synonyms**: Glossaries*

H

Harmonization edit term [451]
In the context of crosswalking, metadata schema of the source and the target standards are represented in the same syntax during harmonization. Related Guide [33] ***Synonyms**: Harmonize*

I

Ingest edit term [452]
In the context of metadata, the ingest process is the method by which metadata is read into a system (e.g. a database import).

M

Machine-Readable edit term [453]
In the context of metadata, formatted in a way that is well defined and processable by the system's software and hardware. Metadata with this characteristic can be discovered, ingested, and presented by an electronic system (also known as 'computable'). Related Guide [21] ***Synonyms**: Machine Readable, Machine Readability*

Metadata edit term [454]
Data about data. Metadata provides a context for research findings, ideally in a machine-readable format. It enables discovery of data via an electronic interface, and correct use and attribution of findings. Related Guide [6]

Metadata Classification edit term [455]
Grouping of metadata values, based on shared criteria. Related Guide [19] ***Synonyms**: Metadata Class, Classes*

Metadata Element edit term [456]
Individual instance of a metadata label and value pair. For example, "creator: John Doe" is a metadata element. Related Guide [118] ***Synonyms**: Metadata Elements, Metadata Fields, Element, Parameter, Metadata Parameter, Field, Metadata Parameters, Metadata Properties, Parameters, Metadata Field, Fields, Elements*

Metadata Extension edit term [457]
Addition to a metadata standard that allows users to provide information in additional fields, or additional ways, that were not mentioned in the original standard. Related Guide [287] ***Synonyms**: Extension, Metadata Extensions, Extensions*

Metadata Instance edit term [458]
A metadata document describing a resource in a standards-compliant manner For example, the Everglades Hydrology and Water Quality Data document [161] provided in XML by the USGS. Also, see other MMI-provided metadata instance examples [25]. ***Synonyms**: Instance*

Metadata Interoperability edit term [459]
The ability of two or more information systems to exchange metadata with minimal loss of information. Related Guide [8] ***Synonyms**: interoperability, interoperable, Metadata Interoperability*

Metadata Label edit term [460]
A descriptor for a metadata value. This can be thought of as a question to which the value is providing an answer. For example, for the metadata label "date", the metadata value could be "March 16, 2008". ***Synonyms**: Label*

Metadata Profile edit term [461]
The community-specific application of a metadata standard. Related Guide [116] ***Synonyms**: Profiles, Metadata Profiles, Profile*

Metadata Specification edit term [462]
Any description of how to store metadata. Specifications have no limitations on the level of required documentation and no requirement for formal approval, publishing or governance by a broad community-based organization. Related Guide [27] *See also: Metadata Standard [157]* ***Synonyms**: Specifications, Metadata Specifications, Specification*

Metadata Standard edit term [463]
A set of documented rules which define the creation of metadata by providing a combination of terminology (vocabularies), syntactical rules, format rules, and other requirements. Metadata standards are approved, published and governed by a formal body or organization with broad community-based representation (international or national). Related Guide [158] *See also: Metadata Specification [104]* ***Synonyms**: Standard, Metadata Standards, Standards*

Metadata Value edit term [464]
Metadata values are the content connected to metadata labels in a metadata element. For example, if the metadata label is "date", the metadata value could be "May 13, 2007". Related Guide [120] ***Synonyms**: Values, Metadata Values, Value*

Multi-Level Vocabulary edit term [465]
A managed list of metadata terms, where the terms are organized into categories. Multi-Level vocabularies include taxonomies and subject headings. Related Guide [252] ***Synonyms**: Multi Level Vocabularies, Multi-Level Vocabularies, Multi Level Vocabulary*

O

Ocean Observing Network edit term [466]
Connected system of data collection nodes. ***Synonyms**: OOS*

Ontology edit term [467]
A type of relational controlled vocabulary, which provides for categories, relationships, rules and axioms among metadata elements. Typically a hierarchy of classes and terms, an ontology is a machine-readable way of relating metadata terminology. Related Guide [51] ***Synonyms**: Ontologies*

P

Protocol edit term [468]
A strategy for transmitting data between systems. A protocol can be used not only over the internet, between computers, but also between applications running anywhere. Examples: FTP, SNMP, SSH. ***Synonyms**: Protocols*

Provenance edit term [469]

The record of how a particular value or record came to be. Provenance can include things like when, by whom, and how the item was created or modified.

***Synonyms**: History*

R

Relational Vocabulary edit term [470]
: Managed list of acceptable terms that makes use of relationships between metadata terms. Relational vocabularies include thesauri, semantic networks and ontologies. Related Guide [237] ***Synonyms**: Relational Vocabularies*

Relationship edit term [471]
: Connections between metadata terms within a vocabulary. These relationships can connect terms by scope, provenance, or other well-defined criteria. ***Synonyms**: Relationships*

S

Scalability edit term [472]
: The ability of a metadata system to expand. Well-designed systems are established with the flexibility to scale up to larger data sets, enhanced metadata requirements, and a variety of growth factors. ***Synonyms**: Scalable*

Schema edit term [473]
: In the context of metadata, a description of the data represented within a database. ***Synonyms**: Schemas*

Semantic Framework edit term [474]
: A semantic framework guides a specific development to make use of computer-interpretable programming languages, such as XML, to create systems which promote and allow semantic interoperability. Both semantic interoperability and the Semantic Web rely on the backbone of a semantic framework. May also refer to the Marine Metadata Interoperability's own Semantic Framework [270]. Related Guide [271]

Semantic Interoperability edit term [475]
: The ability of multiple systems to exchange information in useful ways; in particular, the ability for each system to 'understand' the terms of the other sufficiently to use those terms correctly. Related Guide [57]

Semantic Mapping edit term [476]
: In the context of crosswalking, elements in the source schema are explicitly mapped to elements in the target schema during semantic mapping. Related Guide [34] ***Synonyms**: Semantic Mappings*

Semantic Network edit term [477]
: A type of relational controlled vocabulary consisting of lists of terms/concepts and directed relationships. Related Guide [50]

Semantic Technology edit term [478]
: Semantic technology provides the meaning behind data alongside the data itself. Software written to enable semantic technology explicitly separates the underlying code, data input and output, and data meaning from one another. ***Synonyms**: Semantic Technologies*

Semantic Web edit term [479]
: The transformation of the web from an inherently human-interpretable medium to an inherently computer-interpretable medium. In the semantic web, machines can read and understand the content published in the network.

Subject Heading edit term [480]
: A type of multi-level controlled vocabulary in which metadata values are classified into categories which may be broad classes. Related Guide [48] ***Synonyms**: Subject Headings*

T

Taxonomy edit term [481]
: A multi-level controlled vocabulary in which metadata terms are grouped according to subject-specific classes, usually hierarchical. Related Guide [47] ***Synonyms**: Taxonomies*

Thesaurus edit term [482]
: A type of relational controlled vocabulary which provides a list of terms, with specific relationships between the terms. Related Guide [49] ***Synonyms**: Thesauri*

Transformation edit term [483]
: In the context of crosswalking, transformation is the process of creating a target instance of the metadata description from the source instance. Related Guide [36] ***Synonyms**: Transformations*

U

Usage Vocabulary edit term [484]
: The set of terms used to identify, analyze, or re-use data values in the native form of the data asset. Related Guide [39] ***Synonyms**: Usage Metadata, Usage Vocabularies*

V

Vocabulary edit term [485]
: A set of terms (e.g., words) that are used in a specific community. Related Guide [12] ***Synonyms**: Vocabularies*

Vocabulary Mapping edit term [486]
: Documents that map metadata terms between different controlled vocabularies. Related Guide [54] ***Synonyms**: Vocabulary Mappings*

Vocabulary Term edit term [487]
: A potential metadata value that is part of a set intended to restrict the available options in a particular metadata element. ***Synonyms**: Vocabulary Terms*

W

Web Service edit term [488]

Standardized way of integrating Web-based applications using open standards over an Internet protocol backbone. Web services share business logic, data and processes through a programmatic interface across a network. The applications interface, not the users. ***Synonyms****: Web Services*

Additional Metadata Glossaries on the Internet

- Dublin Core Metadata Initiative (DCMI) Glossary [489]
- Introduction to Metadata, Pathways to Digital Information [490], Available from the Research at the Getty

Glossary of Metadata Acronyms

A glossary of the terms used on the Marine Metadata Interoperability project web site.

This glossary provides metadata-related definitions for the acronyms used on this web site.

Metadata-Related Acronyms

Add term [491] Edit acronym glossary [492] Glossary settings [420]

A

ADN edit term [514]
ADEPT/DLESE/NASA metadata framework

ANSI edit term [515]
American National Standards Institute

API edit term [516]
Application Program Interface

ASCII edit term [517]
American Standard Code for Information Interchange

B

BODC edit term [518]
British Oceanographic Data Center

C

CoRIS edit term [519]
Coral Reef Information System

CSDGM edit term [520]
Content Standard for Digital Geospatial Metadata

CSV edit term [521]
Comma Separated Value

CTD edit term [522]
Conductivity - Temperature - Depth

CV edit term [523]
Controlled Vocabulary

D

DCMI edit term [524]
Dublin Core Metadata Initiative

DIF edit term [525]
Directory Interchange Format

DL edit term [526]
Description Logistics

DLESE edit term [527]
Digital Library for Earth System Education

DMAC edit term [528]
Data Management And Communications

DOI edit term [529]
Digital Object Identifier

DTD edit term [530]
Document Type Definition

E

ER edit term [531]
Entity-Relational (as in Entity-Relational Diagram)

ERESE edit term [532]
Enduring Resources for Earch Science Education

F

FGDC edit term [533]
Federal Geographic Data Committee

G

GCMD edit term [534]
Global Change Master Directory

GIS edit term [535]
Geographic Information System

GML edit term [536]
Geography Markup Language

GUI edit term [537]
Graphical User Interface

H

HTML edit term [538]
HyperText Markup Language

I

IEEE edit term [539]
Institute of Electrical and Electronics Engineers

IOOS edit term [540]
Integrated Ocean Observing System

ISO edit term [541]
International Standards Organization

L

LC edit term [542]
Library of Congress

M

MARC edit term [543]
MAchine Readable Cataloging

MB37 edit term [544]
MultiBeam 37

MBARI edit term [545]
Monterey Bay Aquarium and Research Institution

METS edit term [546]
Metadata Encoding and Transmission Standards

MMI edit term [547]
Marine Metadata Interoperability

MOOS edit term [548]
Monterey Ocean Observing System

MUSE edit term [549]
MOOS Upper-Water-Column Science Experiment

N

NBII edit term [550]
National Biological Information Infrastructure

NetCDF edit term [551]
Network Common Data Format

NISO edit term [552]
National Information Standards organization

NOAA edit term [553]
National Oceanic and Atmospheric Administration

NSDI edit term [554]
National Spatial Digital Infrastructure

NSDL edit term [555]
National Science Digital Library

O

OAI edit term [556]
Open Archives Initiative

OAI - PMH edit term [557]
Open Archives Initiative Protocol for Metadata Harvesting

OCLC edit term [558]
Online Computer Library Center

OGC edit term [559]
Open Geospatial Consortium

OWL edit term [560]
Web Ontology Language

P

PMH edit term [561]
Protocol for Metadata Harvesting

PURL edit term [562]
Persistent Uniform Resource Locator

Q

QA edit term [563]
: Quality Assurance

QC edit term [564]
: Quality Control

R

RDF edit term [565]
: Resource Description Framework

S

SDSC edit term [566]
: San Diego Supercomputer Center

SGML edit term [567]
: Standard Generalized Markup Language

SIO edit term [568]
: Scripps Institution of Oceanography

SKOS edit term [569]
: Simple Knowledge Organization System

SOAP edit term [570]
: Simple Object Access Protocol

SRL edit term [571]
: Semantic Rule Language

SST edit term [572]
: Sea Surface Temperature

SUB edit term [573]
: State and University Library

SVN edit term [574]
: SubVersioN

U

UDDI edit term [575]
: Universal Description ***Synonyms****: Discovery and Integration*

UML edit term [576]
: Unified Modeling Language

URI edit term [577]
: Uniform Resource Identifier

URL edit term [578]
: Uniform Resource Locator

URN edit term [579]
: Uniform Resource Name

USGS edit term [580]
: United States Geological Survey

V

VINE edit term [581]
: Vocabulary INtegration Environment Tool

W

W3C edit term [582]
: World Wide Web Consortium

WFS edit term [583]
: Web Feature Service

WHOI edit term [584]
: Woods Hole Oceanographic Institution

WSDL edit term [585]
: Web Services Description Language

X

XBT edit term [586]
: eXpendable BathyThermograph

XML edit term [587]
: eXtensible Markup Language

XSL edit term [588]
: eXtensible Stylesheet Language

About the Guides Editorial Group

The MMI Guides are being developed through a community-based effort, involving volunteers from many research institutions and agencies across the US and Canada, and supported in part by a grant from the National Science Foundation. The authors of each individual guide are listed on that guide's page. The Editorial Group reviews the guides, ensures quality and consistancy across guides, develops the table of contents and defines needed guides, and seeks external feedback on the guides.

Editorial Group Membership

- Paul Alexander, Marine Metadata Interoperability Project (Consultant)
- Nan Galbraith, Woods Hole Oceanographic Institution [589]
- John Graybeal, Monterey Bay Aquarium Research Institute [590]
- Anthony W. Isenor, Defence Research and Development Canada [591]
- Ellyn T. Montgomery, U.S. Geological Survey [592]
- Caryn Neiswender, Scripps Institution of Oceanography [593]
- Karen Stocks (Chair), San Diego Supercomputer Center, UCSD [594]
- Chris Stuart, San Diego Supercomputer Center, UCSD [594]
- Stephanie Watson, Independent Consultant

Past Editors

The MMI project thanks past members of the Guides Editorial Group for their contributions to the guides: Luis Bermudez, Julie Bosch, Kathryn Joyce, Kyle Hogrefe, Andy Maffei and Steve Miller

In particular, the version of the MMI Guides available in 2007 and before are the product of and should be credited to: John Graybeal, Luis Bermudez, Nan Galbraith, and Stephanie Watson.

Contact us!

If you're interested in learning more about the MMI Guides Editorial Group, our mission or current projects, please contact us [2].

Deprecated Guides

This section contains Guides that have been deprecated. They are here for historical purposes only.
The following guides have been deprecated and are no longer available to the public. Aliases and redirects have been updated to point to current guide documents with analogous content (or removed completely if the content is no longer covered in the guides).

Ontologies {DEPRECATED}

Definition of an ontology (relational vocabulary) in marine metadata.

In computer science an ontology is an explicit and formal specification of mental abstractions, that conforms to a community agreement about a domain and design for a specific purpose (Gruber, 1993). It is different from the term Ontology (first letter in upper case) used in Philosophy to describe the existing things in the world (Fonseca, 2001). Different abstractions, specifications and agreements exist among communities, so different domain ontologies exist, while only a single Ontology is possible. An ontology provides the structure of the controlled vocabulary [595] similar to a dictionary [596] or a thesaurus [597]. The vocabulary agreed to by a community is the expression of concepts (i.e. mental abstractions) of their domain. Since a concept can be expressed in different ways and differ in meaning from one person to another, the controlled vocabulary helps to solve semantic incompatibilities.

- F.T. Fonseca, ONTOLOGY-DRIVEN GEOGRAPHIC INFORMATION SYSTEMS, Doctor of Philosophy, The University of Maine, 2001.
- T. Gruber, A Translation Approach to Portable Ontology Specification., Knowledge Acquisition 5(2), 199-220, 1993.

Ontologies vs. Controlled Vocabularies

A formal specification of a vocabulary can be found as a plain list of words, a dictionary, a taxonomy, an Entity-Relational (ER) diagram, an Object Model in Unified Modeling Language (UML) diagram, an eXtensible Markup Language (XML) schema and possibly many others. What makes a controlled vocabulary an ontology is that in an ontology the concepts are defined explicitly by creating classes [598]. A class is created using a mental abstraction, which can be a classification, an aggregation or a generalization [Batini, 1992]. For example, a list of terms such as USA, Germany, and Colombia do not represent any explicit conceptual relation until an explicit class Country is abstracted to classify them.

In addition to this requirement an ontology needs to conform to strict hierarchical subclass relationships between the classes [Gruber, 1993]. Also, in ontologies the classes have properties and relations [599] among them.

It should be noted that these terms (in particular 'ontology') have been defined many different ways in different publications. Deborah McGuinness, for example, has proposed that an ontology could be construed as including the entire spectrum of controlled vocabularies. We have documented in this FAQ one of the more common discriminators, but other papers and usage may vary.

References:

- C. Batini, S. Ceri and S.B. Navathe, Conceptual Database Design, The Benjamin/Cummings publishing Company, Inc., Redwood City, California, 1992.
- T. Gruber, A Translation Approach to Portable Ontology Specification., Knowledge Acquisition 5(2), 199-220, 1993.

Ontology Tutorials

Tutorials developed by MMI for working with ontologies and ontological tools

- MMI Ontology Creation Guidance [379]
- Using Protege to Edit OWL Ontologies [600]
- Ontology Web Services Tutorial [381]
- URI and Namespace Tutorial [382]

Ontologies as an Interoperability Tool

This guide addresses ontologies as tools for enhancing interoperability between data systems. It provides an overview of ontologies and illustrates how they are used.

Background

There are two major search problems addressed by semantic interoperability [601]between data systems:

1. We cannot find all the data we are seeking.
2. We get too many results and they are difficult to classify.

Ontologies are mechanisms that can be used to help solve these problems. An ontology is a type of controlled vocabulary, which provides for categories, relationships, rules, and axioms among metadata values. Typically a hierarchy of terms, an ontology is a machine-readable way of relating metadata terminology.

Ontologies provide many capabilities. They can satisfy any requirement that a controlled vocabulary can satisfy (although some reformatting might be needed), including providing definitions, controlling possible answers to questions, and ensuring uniform spelling. Because ontologies support a rich set of relationships among different vocabulary terms, they enable a much fuller understanding of terminology and concepts than most controlled vocabularies. For example, the MMI Platforms Ontology (under development) [602] currently includes categorizations of different kinds of oceanographic platforms, as well as complex properties, such as types of mobility, and other platform qualities.

How Ontologies Can Help

Ontologies can be used by automated tools to power advanced services such as more accurate web search, intelligent software agents, and knowledge management. By formalizing relations between concepts of one or more collections in a machine-readable language, ontologies can facilitate interoperability. These concept descriptions determine the format in which the information is kept, and establish the actual conceptual information, or semantic content, that is defined in the ontology. Agreements also should be reached about the community and technical processes used to modify the ontology. Finally, ontologies are designed to be computer-usable (also known as 'computable') - their format and rules are specified so that the information can be found, exchanged, and applied by computer systems, without additional human intervention.

Some examples of how ontologies can faciltate interoperability:

- Knowledge of a Domain
    - This image of a natural catastrophe ontology [603] demonstrates how an ontology can represent a domain of interest (from Robert Laurini [604] INSA –Lyon). Ontologies can fully represent a domain of interest (using concept terms and relationships) and thus enhance interoperability.
- Mappings Between Controlled Vocabularies
    - Controlled vocabularies are important, but there is rarely only one controlled vocabulary relevant to a domain of interest. Different funding sources, project purposes, program histories, etc. lead to different controlled vocabularies for a given domain. Mappings between controlled vocabularies, normalized in ontology representation languages such as the Resource Description Framework [605] or the Web Ontology Language (OWL) [318], can consist of identifying terms in each vocabulary as equivalent to, broader than, narrower than, or a subclass of terms in another vocabulary. Such ontology representations and mappings can enhance interoperability between data systems in that the use of specific search terms is no longer necessary. The mappings between terms in different controlled vocabularies used in different data systems can allow the user to find additional information. For example, at the MMI Advancing Domain Vocabularies Workshop in 2005, we demonstrated the enhanced ability to quickly find sea surface temperature data sources (regardless of whether "SST", "sea surface temperature", "Ocean Temperature" variations were used), using an MMI semantic mediation service called Semor [606]. Semor is a semantic mediation service for earth science terminologies. Terminologies are expressed in ontologies following the RDF model [317]. Users can query terminologies using RDF query languages or simple text matching queries. This service helps users discover what a term means and its relationships to other terms.
- Mappings Between Categories/Hierarchies of Concepts
    - Taxonomies (or other hierarchies) used by different data systems, as well as within a data system, may vary. Ontologies, and mappings between ontologies, can facilitate interoperability between these higher level categorizations. For example, the Oregon Coastal Atlas and the Marine Irish Digital Atlas, which strive to interoperate as components of an International Coastal Atlas Network, use different classifications for grouping their mapping data sets to help users find data sets of interest. MMI is working with this group to create an interoperability prototype [321] between the two atlases, using an upper ontology, as well as mappings between classifications and terms.

How Ontologies Work

Ontologies can represent concepts (as classes), individuals (members or instances of the classes), characteristics of each concept (as properties), and relationships between the concepts in a machine-usable language, based on the Resource Description Framework [317]. RDF is a graph data model, where concepts are represented by nodes, and the relationships between them represented by the lines linking them. RDF, which employs subject-property-object triples and Uniform Resource Identifiers to define ontologies, makes it possible for computers to readily use the information directly represented in ontologies. In addition, depending on the precise type of relationships allowed in a given ontology, computers can 'reason' in various ways -- drawing real-time conclusions relating to terms and data -- using the knowledge embodied in the relationships of the ontology. The standard framework that ontologies provide for representing information and relationships enable many general-purpose software capabilities that would not be feasible with other technologies. For example, search engines can 'understand' that a person looking for 'coastline' is also interested in 'shoreline' within physical coordinates that correspond to an oceanographic geospatial location. Training and testing tools can be written that leverage the information in an ontology, without changing the tool, to present new information to a student. And as 'reasoning systems' become more advanced, the 'raw knowledge' in the ontology can be leveraged with other systems and ontologies, giving computer systems a much more general ability to deal with environmental concepts.

A controlled vocabulary in a simple format, such as ASCII, can be converted into RDF (by using the MMI tool Voc2RDF [281], for example). The terms in this new ontology can then be mapped to terms from another ontology that covers the same subject matter. Tools exist that facilitate mappings between vocabularies that are formatted as RDF or OWL files, and save the result in a similar format. (For example, one free tool is the MMI tool, VINE [290].) The resulting OWL-formatted mappings file can then be used in web services to search (or otherwise interoperate between) multiple data systems for the domain of interest. This advanced search/interoperability is founded on the computers' new "understanding" of the meanings of terms and the relationships between them across the different data systems.

Have a specific question about ontologies? Ask MMI! [299]

Ontologies as an Organizational Aid

How ontologies can help you organize information

Ontologies can help organize information about a topic domain, either within a data system or across different data systems. They can achieve this because of the basic features of ontologies. These features:

- Provide a means for defining concepts so that they can be consistently accessed by a computer.
  - For example, MMI is developing a platforms ontology, which defines different types of oceanographic platforms and their properties (e.g., mobile or not), so that searches for data sets by platform type, or by properties of a platform, can be done automatically across data systems, regardless of the specific terminology used.
- Provide hierarchical frameworks for organizing concepts.
  - Classes, subclasses, and “is a part of” are examples of the hierarchical organizational framework provided by ontologies for concepts. In the MMI platforms ontology, the superclass concept is "Platform", with its subclass concepts, "AirAndOuterSpaceBasedPlatform", "EarthBasedPlatform", and "WaterBasedPlatform".
- Permit the articulation and accessibility of relationships between concepts (instances) and properties.
  - These relationships can then be used automatically by computers to infer additional associations between concepts. Ontologies can automatically classify concepts according to the properties defined for those concepts. For example, a water-based sensor could be classified as a subclass of a sensor, simply if a property of the water-based sensor is also a property of a sensor. In other words, the property can be used to classify water-based sensors as a type of sensor, without having to manually create water-based sensors as a subclass of sensors.
- Can be extended to provide any kind of relationship, or mapping, between individual terms in separate ontologies.
  - For example, at the MMI workshop, Advancing Domain Vocabularies [607], a sensors working group identified and mapped sensor-related terms from several vocabularies, including those from WHOI, MBARI, LDEO, SIO, TAMU, NGDC, CO-OPS, ACT, and BODC. This work was the precursor to the development of the MMI Sensor Ontology project [608], the goal of which is to develop a sensor ontology, based on existing vocabularies and the mappings initiated at the MMI workshop.

Through application of these features, ontologies provide a variety of higher level knowledge-management capabilities, each of which helps organize information and knowledge. Ontology applications:

- Provide consistency to the terms used in metadata records.
  - Consistency in metadata is essential to keeping information organized, making it discoverable, and enables interoperability between data systems.
- Can be used to generate knowledge bases about one or more specific domains(s).
  - Each ontology represents a set of knowledge about some topic area. By connecting related ontologies, the knowledge framework can be extended to cover a wider domain.
- Provide more powerful terms for filling out metadata records, so that they can better represent information.
  - By formally defining the terms used in metadata records, and enabling those terms to be mapped to other terms relevant to that community, ontologies extend the completeness, precision, computability, and extensibility of the metadata records. (Each term in an ontology can carry with it the context of the entire model, due to more complex semantic statements and the inference capabilities of Description Logics). For example, ontologies can define the values used to complete metadata fields like Keywords. The terms can then be mapped to other vocabularies, and interoperability is facilitated between different metadata records, regardless of the specific terminology used.
- Provide greater descriptive detail for metadata models and metadata specifications, so that they can more clearly capture and organize information.
  - By formally defining the field names used in metadata standards, and enabling those terms to be mapped to other terms in other standards, ontologies create a more precise and interoperable standards framework. For example, ontologies can organize and characterize the terms in different metadata standards (e.g., "creator" vs. "author"), enabling both better understanding by people using the standard, and greater interoperability between different standards.
- Support discovery and understanding through interactive navigation.
  - The relationships captured by ontologies are analogous to those in topic maps: they tie together the different terms of the ontology. Visual presentation of these relationships provides a different way to view the knowledge model represented by the ontology, and interactive tools allow it to be easily explored, often with serendipitous results. These explorations can be built in to the interfaces used to discover data sets and other scientific materials, allowing searches to be qualified in ways that make sense for the particular subject domain.

Have a specific question about ontologies? Ask MMI! [299]

Software Tools for Ontologies

A list of widely-used and well-regarded software tools for ontologies.

Software Tools for Ontologies

Below, several types of software tools for ontologies are described. At present, no single tool can provide all the functions listed below. Therefore, it is necessary to use a combination of tools for most tasks. These tools are a good starting point for those seeking more information about ontology tools. This is not an exhaustive list, but presents a selection of tools that are relatively widely-used and well-regarded in the ontology community.

Ontology tools are available for a wide variety of functions, including conceptual modeling (as the first step of ontology development), designing and editing an ontology, visualizing an ontology, transforming or converting an ontology, mapping between ontologies, and merging multiple ontologies into a single, coherent ontology. Domain-specific ontology tools are also available.

Conceptual Modeling

These tools do conceptual modeling – a first step toward building an ontology. Conceptual modeling is the process of identifying concepts (typically, nouns) and their relationships (typically, verbs). These concepts and relationships can then be formalized using an ontology design and editing tool.

Concept Maps (CMAP) Tools [609]
CMAP tools are free to educational and government agencies, as well as individuals. Among other functions, the software allows users to easily create knowledge models called concept maps. Users identify important concepts (typically, nouns) in their domain, as well as relationships (typically, verbs) between concepts.

Maryland Information and Networks Dynamic lab Semantic Web Agents Project [610] (MINDSWAP)
MINDSWAP, among other functions, offers a variety of software, including conceptual visualization tools (to view concepts and relationships in graphical form) and Pellet (a Descriptions Logic Reasoner, which can automatically make inferences about concepts).

Design and Edit

These tools allow for the development of ontologies in an ontology representation language (such as RDF [317] or OWL) [318] and editing existing ontologies. They can vary in terms of their usability, suitability to ontology representation languages, capabilities to develop full ontologies, as well as abilities to align (identify semantic equivalents across) domain and upper ontologies. An upper ontology is a basic ontology that contains the fundamental concepts and relationships across domains that are required to understand other concepts. The following list is a sample of ontology design and editing tools:

Protégé [611]
Protégé is one of the most widely used ontology development and editing tools. It was developed at Stanford Medical Informatics. Protege is open source, free, has a supportive community, has been used across widely varying domains, has extensive plug-ins available, and offers an associated on-line ontology library. A complementary reasoner (a piece of software that examines an ontology and automatically infers classifications that have not been explicitly stated by an ontology developer), like FaCT [612] or PELLET [613], is required for inferencing.

Topbraid Composer [614]
Topbraid Composer is a modeling environment for developing Semantic Web ontologies (in RDF [317]) and building semantic applications. It can be downloaded freely and includes a built-in reasoner. Topbraid may be used with versioning tools, such as Subversion (SVN), to provide a collaborative ontology development environment.

Visualize

Visualization tools allow the user to see an ontology in a graphical representation. This can make it easier to understand an ontology than simply viewing it as an OWL file or as a tree in an ontology editor.

IsaViz [615]
IsaViz is a visual environment for browsing and authoring RDF models represented as graphs.

Ontology Graph [616] (Ograph)
Ograph provides a graphical representation of DARPA Agent Markup Language [617] (DAML) + Ontology Inference Layer [618] (OIL) (DAML + OIL was a precursor to OWL as an ontology representation language) and OWL ontologies.

Transform

These tools convert files in basic formats that are commonly used in the sciences (e.g., ASCII) to ontology representation languages, such as RDF or OWL. This conversion (or harmonization) step is usually a necessary precursor to mapping between vocabularies or ontologies.

Voc2OWL [619] and Voc2RDF [281]
Voc2OWL and Voc2RDF are MMI-developed tools that convert ASCII files to OWL or RDF. They are free and open source.

Map and Merge

Ontology mapping tools provide the ability to map between the terms in two or more vocabularies/ontologies. Mapping facilitates interoperability among data systems that use controlled vocabularies. Merging tools help make two ontologies into a single coherent ontology.
Vocabulary Integration Environment (VINE) [290]
VINE is an MMI-developed tool that allows mappings among terms used in different controlled vocabularies. The mappings are in the form of three different relationships (broader than, narrower than, and same as). The input to VINE is any OWL file in XML format. The output is a text file, which can be used by web services to automatically search different data systems (regardless of the specific search terms used.)

Ontomerge [620]
Ontomerge was developed to merge ontologies. Ontomerge takes the union of the axioms (statements in an ontology that serve as starting points from which other statements in an ontology are logically deduced or inferred) defining the ontologies, using XML namespaces. Ontomerge adds bridging axioms to relate the terms in one ontology to the terms in the other. Inferences can be conducted in this merged ontology.

Domain-Specific Ontology Tools

Some tools have been developed specifically for particular domains. Although they were developed for specific domains, it may be possible to use these tools as models, and adapt them for other domains.
Gene Ontology Tools [621]
The Gene Ontology (GO) project (of the Gene Ontology Consortium, a set of model organism and protein databases and biological research communities actively involved in the development and application of the Gene Ontology) provides a controlled vocabulary to describe gene and gene product attributes in any organism. The project also includes ontology files, annotations, the GO database of ontologies and annotations related to the GO, and tools for using the GO, including, for example:
OBO-Edit downloads - an open source, platform-independent application for viewing and editing their special OBO flat-file format ontologies);
AmiGO - an interface to search and browse the ontology and annotation data provided by the GO consortium; and
GO Online SQL Environment - to directly query the GO Database. Online

Additional References

A survey of ontology editors [622], including a table outlining the different ontology editors surveyed

A more comprehensive listing [623] of available ontology tools [624]

Tool evaluations by MMI [383] (underway as of January '08)
[299]

MMI Guides Working Drafts {DEPRECATED}

GUIDES WORKING DRAFTS HAVE BEEN DEPRECATED

If you need to move content from here because it is being readied for publication, please follow the instructions for adding new Guides content [625].

This is the book that contains working draft book pages for the Guides book. To create a draft, you may add a child page to one of the main pages, or create a dummy hierarchy to show where the page or pages will go.
After completing your page(s) and getting approval from the team leader, simply change the Parent of your page to the appropriate place in the real Guides.

About the MMI Guides Editorial Group

About the MMI Guides Team

The MMI Guides are being developed through a community-based effort, involving volunteers from many research institutions and agencies across the US and Canada, and supported in part by a grant from the National Science Foundation.

The Guides Team's mission is to help researchers and data managers follow best practices in metadata development and distribution, and to foster community involvement in this process. To reach our goals, we have distilled the more complex aspects of marine metadata into easily accessible documents, and developed a learning environment which we hope will facilitate both technical understanding and an appreciation of the role and importance of marine metadata.

Editorial Group Membership

- Paul Alexander, Monterey Bay Aquarium Research Institute
- Nan Galbraith, Woods Hole Oceanographic Institution
- Andrew Gale, San Diego Supercomputer Center, University of California San Diego
- John Graybeal, Monterey Bay Aquarium Research Institute
- Anthony W. Isenor, Defense Research and Development Canada
- Ellyn Montgomery, U.S. Geological Survey
- Caryn Neiswender, Scripps Institution of Oceanography
- Karen Stocks (Chair), San Diego Supercomputer Center, University of California San Diego

Past Editors

The MMI project thanks past members of the Guides Editorial Group for their contributions to the guides: Luis Bermudez, Julie Bosch, Kathryn Joyce, Kyle Hogrefe, Andy Maffei and Steve Miller

An earlier version of the MMI Guides, available before 2008, was developed by, and should be credited to: John Graybeal, Luis Bermudez, Nan Galbraith, and Stephanie Watson.

Contact us

If you're interested in learning more about the MMI Guides Editorial Group, our mission or current projects, please contact us [2]!

Brain Dump

This is a place to unload ideas, writings or ramblings that don't fit elsewhere at the moment...

****

Need and intermediate-level ontology guide, to bridge the intro material and Stephanie's advanced guide.

****

From John, email 7/1/2008:

Peter Fox wrote up the following semantic technologies as being applicable to any current semantic interoperability effort. Seems to me most of these should end up in a Guide somewhere or another, sooner or later (and I know many of them are, i just wanted to capture this as it went by):

SPARQL

SKOS

RDF

OWL-DL

Protege (OWL-API)

Jena

Pellet

Proof Markup Language (PML)

Eclipse/Spring

IW (Information Web?)

Probe-It!

Sesame2

Just a data point to capture this info. It's nice to have a place to send this kind of data.

****

How to find ontologies? From email from Brand Niemann via John G, 7/25/2008

there are simple ways of discovering ontologies using Google search (*.owl) and tools like Swoogle that we have talked about in our past SICoP Conferences - so also just Google SICoP.

I have Google SICoP and a list of ontologies does not exactly jump out at me. The group is nominally found here: http://ontolog.cim3.net/cgi-bin/wiki.pl?SICoP [626] but I didn't successfully obtain any new ontologies by brief browsing around.

This is a good brief topic for (or within) a guide, perhaps.

****

Units

****

From Ellyn, email 4/16/2007

In thinking about what writing tasks I could sign up for, I realized
that I didn't know whether the CF (climate format) was a standard,

controlled vocabulary or some other entity. Unidata's site says CF is an extension of COARDS, and that both are netCDF conventions, but where does that fit into the framework we've established for the guides? For the groups I interact with, CF is fairly widely used, and I think it might be a good candidate for a case study, but in what category. I know netCDF can be pretty daunting, but it provides great capability for integrating metadata with the data, and there are a lot of tools available for working with it. Is there a non-netcdf based example with a similar widespread user-base that we could also include?

Changing subjects, there are separate topics for "choosing a standard" and "choosing a controlled vocabulary"- it seems that these probably have a lot of overlap, and perhaps a topic like "what do I need to think about to decide what to implement?" would be more helpful. The topics I'm aware of that are important are these: (please add to or ammend the list as needed)
  who funded the project and what are their data sharing requirements?
  who are my likely users?
  are there applications that *must* be able to access the data?
  is international access important?
  what kinds of data do I have? (Images, maps, timeseries measurements, satellite data...)
  what are the widely used standards/vocabularies in my field (seafloor mapping, physical oceanography)?
     is one more mature than the others?
     are any of the above commonly used in related fields (meteorology, atmospheric chemistry)?
  what tools exist to help with implementation?
  what tools exist to facilitate web access?

II.C.1. Standards Bodies

II.C.2. Standards Bodies (ISO/NISO, DCMI, FGDC, OGC, CDMC, Lib of Congress, OAI, ADN)
Author: Karen

II.E.1. Harmonization

Author: ???

II.E.3. Rules

Author?

II.E.4. Transformation

II.E.4. Transformation
Author: ???

III.C.2. Using Controlled Vocabularies in Your System

Use Cases

There are many potential applications of core vocabularies in a data system. We briefly mention a few here; others are represented by our vocabulary use cases [627].

On the ***data search*** side, you might want your system users to have greater chances of finding data sets in external repositories. You know what the GCMD keywords and CF variable names are, but your users may need a simpler vocabulary, or need to be able to refine a search based on those vocabularies. To do this, some technique to discover related terms is necessary. This will depend on a collection of relationships, in which scientists or developers associate one set of terms with another, and a way to use those relationships 'on the fly' in your application or web interface.

On the other hand, you may be providing data archive, and asking users to ***document their data*** in ways that other users can understand. It will be so much more interoperable if your users can use a standard set of community terms that are updated regularly in your interface as the community continues to improve its vocabulary. You want to provide users with easy access to those terms.

For a third and final example, you might want to create a display that ***integrates data*** into a single view, but needs to do so according to a common set of terms (for example, the IOOS Core Vocabulary). You'd like to be able to provide a common list of terms to the users, but then have those terms map to any terms that correspond, taking into account everyone's different vocabularies and variable naming.

Each of these use cases, and many more, can be accommodated by a set of capabilities that you can integrate into your system. Some of these situations have semantically aware examples already developed, while others are in development. We will give an outline of how these systems can work.

You will notice that we primarily talk about simple vocabularies and term mappings in these use cases, but many of these principles also work with more complicated domain descriptions that are contained in 'upper ontologies'.

Core Capabilities

The core capabilities that you will need to access to implement solutions for these use cases include:

A. Finding terms in one or more community or shared vocabularies.
B. Obtaining a list of all the terms in a shared vocabulary.
   This can be as easy as specifying the web link (URL) in your software, with an Accept: header of rdf+xml; or it can be done.
C. Finding terms that are related to your terms (for example, that have narrower meaning, or that have been associated as being similar to the ones you seek).
D. Given a term in one vocabulary, and the name of a second vocabulary, obtain the corresponding term (if any) in the second vocabulary.

Other capabilities will also prove useful, but these are sufficient to discuss the use cases above.

Solution/Approach

We will not provide examples with specific interfaces, as these depend on the services from which you obtain your vocabularies. Instead, we provide descriptions of the general process in each case, and refer you to the documentation for the ontology services that you are using.

A. Data Search

The data search case can rely on a pre-existing set of mappings between your user's terms, or offer the user a choice of terms from existing vocabularies, or search for their terms in the existing vocabularies. To take one example, let's say your user is a modeler working primarily with terms from the CF (COARDS CF Climate and Forecast Standard Names) vocabulary. You want to offer the ability to search GCMD (NASA Global Change Master Directory) for related data sets, but the CF terms are not in the GCMD clearinghouse.

You can provide your user a list of CF terms to choose from by requesting the most recent version of the CF vocabulary (B above). This will download the latest CF vocabulary to your software, and you can easily parse it into constituent terms.

As a second step, once the user has selected one or more terms, you can query the repository directly for related terms in the GCMD vocabulary (D). (The relationships were previously added in this case, provided by previous users who had a similar need for the relationships.) These terms you obtain can then be directly used in a search of the GCMD repository, to find the data sets that match your user's interests.

If you wanted to search more broadly, not just in GCMD but maybe across the internet, you might be interested in *all* the related terms from widely used community vocabularies. This is a similar request (C), but without specifying a target destination vocabulary. Again, relationships added by other users would inform the associations that you obtain.

B. Data Documentation

If you are trying to encourage standards-based documentation, you may want to guide or control the user's entry of descriptive terms. This eliminates misspellings and increases re-usability of user responses, since all the data can be searched using the same vocabulary terms used to document it.

To do so, you will have to download the 'legal' vocabularies as you did in the previous case. But now, the user may need help finding a term. You may want to provide auto-completion, which will just use the terms themselves. More often, it will be helpful to provide a search capability, that lets the user enter a string and you can find matching terms or definitions in the legal community vocabularies.

One option for this, if you have a small set of vocabularies, is to read them into local data storage and work with them there to perform searches. This is also a potential caching strategy. But if there are a lot of vocabularies and you want optimized searching, you can submit the query to the repository (A), which will respond with matching concepts. You can then present the returned concepts to your user for selecting the final terms for documenting his or her data.

C. Data Integration

/ describe what we do in oostethys, or will do /

More Details

More details and considerations for many of these use cases are described as part of MMI's semantic framework concept document, in particular in the Concept of Operations [628].

New Ontologies Section Draft

Ontology in Practice

The following section provides an overview of an actual ontology file including details for the XML code used to produce and store ontological information. We recommend you read this section before proceeding in the guide, even if some of the terms aren't immediately recognizable.

The following section provides an overview of terminology, explaining and illustrating differences in technical terms that are often used together and mean similar things. Then it provides a walk-through of some of the syntax used to construct ontologies, specifically with a much-simplified version of the Wine ontology created by Stanford and modified by the W3C. The full version [629] contains more classes, more properties associated with each class, and more individuals (also called 'terms'). The ontology contains knowledge about the domain 'wine', including relationships to external ontologies and definitions of resources within the wine ontology. You can view a demonstration application [630] that makes use of the Wine Ontology provided by Stanford's Knowledge System, AI Laboratory.

A Note About Resources

At the heart of semantic web technology are methods by which we can describe 'resources' in the world, including properties that resources may have and relationships between resources. Resources can be thought of as things you would want to define, describe, or reference, for example a concept you use, a term, a web page, or a type of relationship between things. Ontologies are made up of resources, each of which has a Uniform Resource Identifier (URI), which are defined and described using XML, RDF, and OWL.

URI vs URL vs URN

When discussing these addresses, there are three acronyms commonly used, and just as often confused: URI, URL, and URN. Before proceeding, it is recommended that you understand the distinction between these three terms.

Simply put, a URI (Uniform Resource Identifier) is a method to identify or name a resource. URLs (Uniform Resource Locators) and URNs (Uniform Resource Names) are both types of URIs. URLs indicate the location of the resource and a way to act upon that resource (the 'http' in http://example.org [631] indicates a transfer protocol). URNs identify a resource and provide it with a name. These two types of URI complement one another and should be used in concert. The full complexities of URIs are well beyond the scope of this document, so for more information, you can start at Wikipedia [632] and move to the more technical w3c [633].

XML, RDF, OWL

Extensible Markup Language (XML), Resource Description Framework (RDF), and Web Ontology Language (OWL) are all markup languages used in computer science for describing things in a document. XML was created to provide a framework which can be used to create syntax for other languages according to agreed-upon standards. RDF is a language used to describe resources on the internet and it uses an XML syntax to do so. This syntax is referred to as XML/RDF. OWL is yet another technology layered on top of RDF/XML. Like RDF, OWL can use XML in constructing documents using an OWL-specific syntax. OWL is generally used for more abstract, higher-level data modeling than XML but uses datatypes directly from XML, such as *double*, *date*, and *float*.

OWL and Object Oriented Programming

Those familiar with Object Oriented Programming (OOP) might recognize some similarities between OWL and concepts found in OOP, specifically classes and instances. While OWL leverages parts of the class/object model, it is only a declarative and logical language rather than an operational one, like C++ or Java. This means you can't perform computations using just OWL and there are no functions, methods, or other means of operating with the language. Also, OWL does not make an assumption that it has data completeness when it comes to a particular object, unlike OOP environments where everything that can be known about an object is contained in the object's type. However, there are projects, like RDFReactor [374], that try to provide OOP proxies for working with ontology data.

Subject, Predicate, Object and the Resulting Triples

RDF is designed to describe resources using a distinct but simple model, involving statements that use subjects, predicates, and objects. These statements take something we would say in English, for example, "The **lightbulb** has a **creator** whose value is **Thomas Edison**," and turn it into something a machine can read. Example 1.1 explores this concept using URLs as the mechanism for identifying the resources. The resulting subject, predicate, object statements are referred to as triples and can also be written using triple notation, also included in Example 1.1.

```
Example 1.1: Subject, Predicate, Object
Assuming we have an ontology describing household objects and another describing inventors.

Subject: http://example.org/householdObjects#Lightbulb
Predicate: http://purl.org/dc/elements/1.1/creator
Object: http://example.org/inventors#ThomasEdison

Triple Notation for this Statement:
<http://example.org/householdObjects#Lightbulb> <http://purl.org/dc/elements/1.1/creator> <http://example.org/inventors#T
```

Defining an Ontology

The first line in Example 1.2 is the very beginning of the ontology. It includes a URI in the form of a URL that identifies the location of the ontology.

```
Example 1.2: Base Referencing

<rdf:RDF xml:base="http://www.w3.org/TR/2003/PR-owl-guide-20031209/wine#">
```

The URL here becomes a base reference for other resources in the ontology that we'll be defining. It's a kind of shortcut that allows us to define further elements without typing the entire URL every time and it works when we are defining elements using rdf:about, rdf:resource, rdf:ID and rdf:datatype. In Example 1.3 we see an OWL class defined with an rdf:ID set to "Wine". This indicates that the resource is located at http://www.w3.org/TR/2003/PR-owl-guide-20031209/wine#Wine, because the rdf:ID is relative to the XML base URL.

```
Example 1.3: Relative to Base

<owl:Class rdf:ID="Wine">
```

Next comes a section devoted to describing the ontology using OWL and RDF syntax, the start of which is illustrated in Example 1.4. This section will generally contain versioning information, comments, and other ontologies that are included in the ontology we're creating.

The first line defines what the section is describing, namely that we're working with an owl:Ontology and that the owl:Ontology information is about a resource at *blank*, which is a roundabout way of saying that we're describing the ontology itself. Because rdf:about makes use of the xml:base shortcut from above and contains nothing, we know that we're discussing the actual base resource, which is our ontology itself.

```
Example 1.4: Ontology Definition

<owl:Ontology rdf:about="">
```

The following five lines contain comments about the ontology, which are simply used for documentation and to provide a human-readable description of a particular resource. You can see here that the information contains some data about data provenance and a basic description of the ontology. You'll also notice that these comments are bracketed on both sides by rdfs:comment tags.

```
Example 1.5: Ontology Commenting

<rdfs:comment>An example OWL ontology</rdfs:comment>
<rdfs:comment>
    Derived from the DAML Wine ontology at
    "http://ontolingua.stanford.edu/doc/chimaera/ontologies/wines.daml"
    Substantially changed, in particular the Region based relations.
</rdfs:comment>
```

XML uses these brackets to indicate the opening () and closing () of sections, much in the same way that HTML uses opening and closing tags to start and stop the italicizing of text or to demarcate a link in text, demonstrated in Example 1.6.

```
Example 1.6: HTML Tag Examples

HTML Examples of opening and closing brackets
Italicizing: <i>Italicized text</i>
Link: <a href="http://example.com">Link text</a>
```

The ontology description section ends with a human-readable label of the ontology, we can see this in Example 1.7 defined as "Wine Ontology". The label is nothing more than a name for the ontology that doesn't use a URI as part of its definition. The next line closes the ontology description section with an owl:Ontology close bracket.

```
Example 1.7: Ontology Label

<rdfs:label>Wine Ontology</rdfs:label>
</owl:Ontology>
```

Classes

Once an ontology has been defined and described, we can add some substantive content in the form of classes. Classes are defined with more OWL tags that indicate what they are (owl:Class) and TODO good description of rdf:ID and reifying RDF resources. In Example 1.8, we can see the syntax that defines the class "Wine". The class description is included in the tag. We can also see that the Wine class is a subclass of PotableLiquid, a class that is defined in the "food" ontology that is hosted by the W3C.

```
Example 1.8: Class Definitions and Description

<owl:Class rdf:ID="Wine">
    <rdfs:subClassOf rdf:resource="http://www.w3.org/TR/2003/PR-owl-guide-20031209/food#PotableLiquid"/><rdfs:subClassOf>
</owl:Class>
```

Class Restrictions

Classes can be defined with a required structure, so that any time we create an instance of the class we know what data is required at a minimum for the individual of the class to be valid. This is akin to having an online form that has required fields that you must fill out before submitting it for processing. It also provides a base for inferencing and reasoning, which are discussed later in the document. In Example 1.9, we can see that there is a restriction on our Wine class, basically that an individual of the Wine class must also have the property "hasMaker" defined with at least one "maker", where maker is equivalent to the winery or vintner of the wine. The tag defines the resource that must be included in an individual, and the tag defines how many of these resources are required per individual. The Wine class includes more restrictions, including that each individual of the Wine class must have at least one "flavor", one "color", and one "locatedIn" property defined.

```
Example 1.9: Class Restrictions

<rdfs:subClassOf>
     <owl:Restriction>
          <owl:onProperty rdf:resource="#hasMaker"/>
          <owl:cardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#nonNegativeInteger">1</owl:cardinality>
     </owl:Restriction>
</rdfs:subClassOf>
```

Language Support

XML has multi-lingual support and RDF, and consequently OWL, takes advantage of this when defining classes. Earlier, we saw an example of how human-readable labels were applied to resources using the tag. These tags can also include translations and language-specific information, as seen in Example 1.10.

```
Example 1.10: Language Support

<rdfs:label xml:lang="en">wine</rdfs:label>
<rdfs:label xml:lang="fr">vin</rdfs:label>
```

Individuals of Classes

Depending on how a particular ontology is structured, it may or may not contain individuals relating to the classes that it defines. The Wine ontology happens to include individuals, so we can examine them in the same context we've been working. In Example 1.11, we can see the resource SantaCruzMountainVineyardCabernetSauvignon, which is an individual of the CabernetSauvignon class. The CabernetSauvignon class has restrictions for hasColor, hasSugar, hasFlavor, and hasBody, all of which are defined for our individual in the example. We also know where this resource exists as the rdf:ID is relative to the ontology's base URL.

```
Example 1.11: SantaCruzMountainVineyardCabernetSauvignon Individual

<CabernetSauvignon rdf:ID="SantaCruzMountainVineyardCabernetSauvignon">
    <locatedIn rdf:resource="#SantaCruzMountainsRegion"/>
    <hasMaker rdf:resource="#SantaCruzMountainVineyard"/>
    <hasSugar rdf:resource="#Dry"/>
    <hasFlavor rdf:resource="#Strong"/>
    <hasBody rdf:resource="#Full"/>
</CabernetSauvignon>
```

We discussed earlier that OWL is generally used for describing abstract, high-level concepts. However, at the individual level we can see some pretty concrete facts about a particular wine, it's winery, the region it's from, and it's flavor profile, which are all features we expect to be knowable about a particular wine. You could make a similar real-world distinction between a sandy beach and a grain of sand, where the sand was a class that we know has certain restrictions on it's definitions. Some of these might include that the sand's individual members must all be small, round, and broken down from larger fragments of rock. If we were to pick up an individual grain of sand and provide it a name, we could then characterize it using properties we have come to expect from individuals of the sand class. For the Wine Ontology, the ontology creator has decided that their ontology's scope will include this type of information. Other ontologies may only include abstract class definitions. Both models are valid ontologies and may have similar or varying uses depending on the application they are intended for.

Inferencing and Reasoning

One of the most powerful benefits of using ontologies is their ability to enable machine-based reasoning over their data. As a basic example, we can examine Stanford KSL's Wine Agent [630]. The application uses the Wine Ontology and other ontologies that describe food definitions and pairing suggestions to match a particular food with a particular wine. This is a rather mundane and simple example of the use of ontologies but will be familiar because of our exploration of the Wine Ontology.

A more advanced example of reasoning can be seen by performing inferencing on the wine ontology to create connections between individuals and classes that the person who created the ontology did not explicitly define. OWL reasonsers are programs written to take advantage of the rules used when constructing OWL ontologies. As an example, look at Images 1.1 and 1.2 below. You will see on first glance that Image 1.2 has many relationships which don't appear in the original version of the ontology. These relationships have been inferred based on the information that was provided when the ontology was created. TODO add example of inferencing.

TODO include graphics showing the flat wine ontology and the reasoned wine ontology with all of the relationships exposed, once I get protege working with GraphViz.

Wine Ontology Visualized: Asserted Model

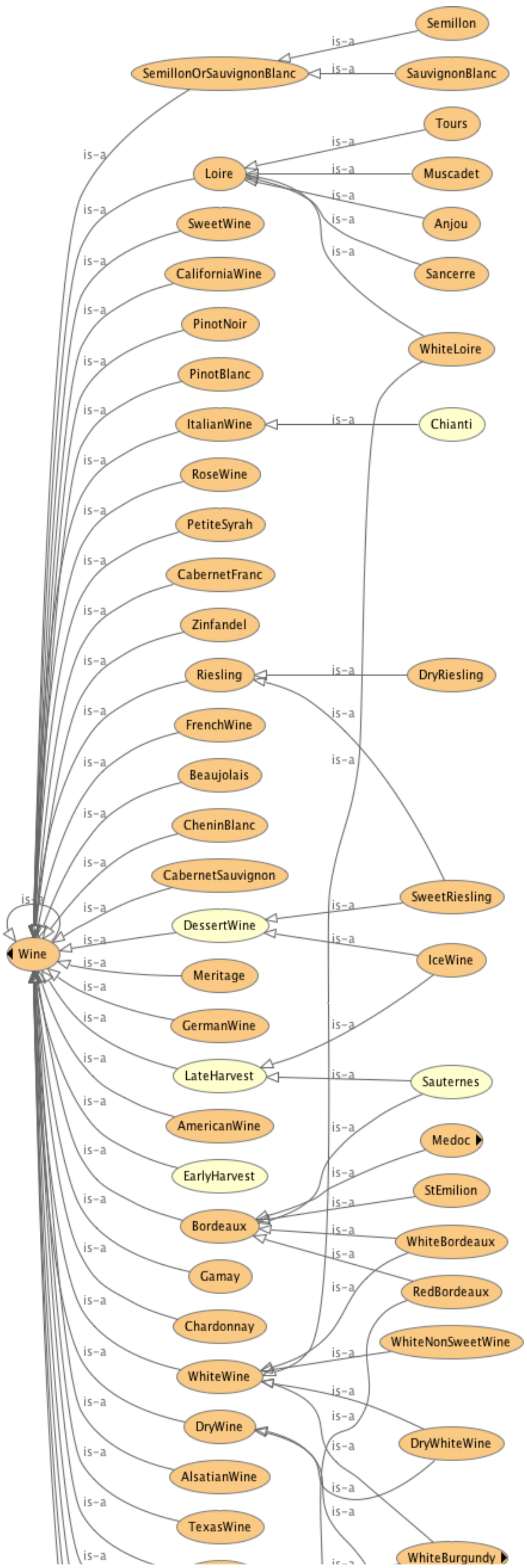
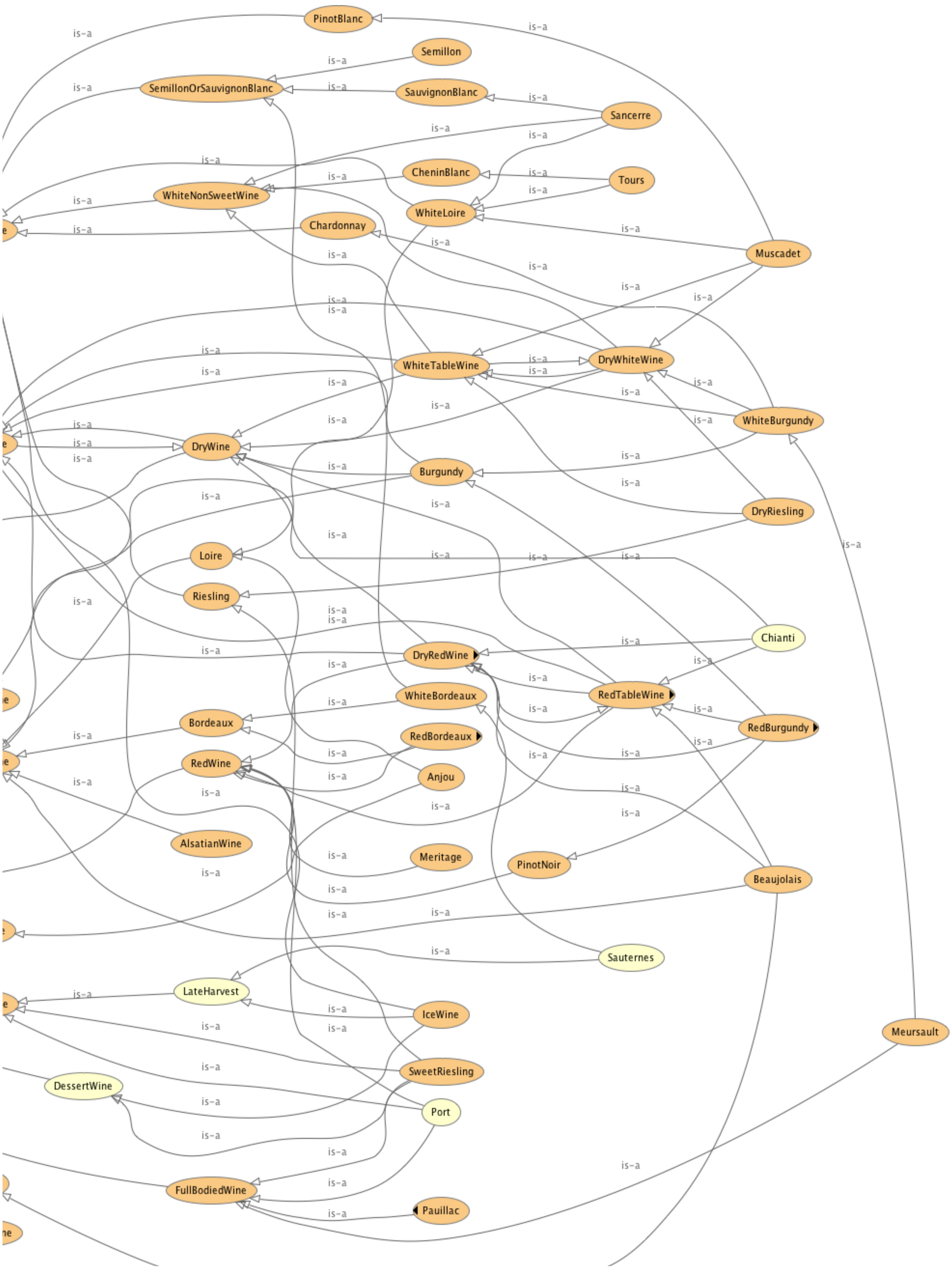

Semillon
SemillonOrSauvignonBlanc
SauvignonBlanc
Tours
Loire
Muscadet
Anjou
Sancerre
SweetWine
CaliforniaWine
PinotNoir
WhiteLoire
PinotBlanc
ItalianWine
Chianti
RoseWine
PetiteSyrah
CabernetFranc
Zinfandel
Riesling
DryRiesling
FrenchWine
Beaujolais
CheninBlanc
CabernetSauvignon
SweetRiesling
DessertWine
Wine
IceWine
Meritage
GermanWine
LateHarvest
Sauternes
AmericanWine
Medoc
EarlyHarvest
StEmilion
Bordeaux
WhiteBordeaux
Gamay
RedBordeaux
Chardonnay
WhiteNonSweetWine
WhiteWine
DryWine
DryWhiteWine
AlsatianWine
TexasWine
WhiteBurgundy
is-a

Image 1.1

ıalized: Inferred Model


PinotBlanc
Semillon
SemillonOrSauvignonBlanc
SauvignonBlanc
Sancerre
CheninBlanc
Tours
WhiteNonSweetWine
WhiteLoire
Chardonnay
Muscadet
WhiteTableWine
DryWhiteWine
WhiteBurgundy
DryWine
Burgundy
DryRiesling
Loire
Riesling
Chianti
DryRedWine
WhiteBordeaux
RedTableWine
Bordeaux
RedBordeaux
RedBurgundy
RedWine
Anjou
AlsatianWine
Meritage
PinotNoir
Beaujolais
Sauternes
LateHarvest
IceWine
Meursault
SweetRiesling
DessertWine
Port
FullBodiedWine
Pauillac
is-a

Vocabularies Re-Org

Working with ontologies

Working with ontologies

Do you need to develop or create your own ontology? If you do, then there are a few things you should be thinking about. Some of these things are technical, and some related to software. Still other things to consider will be related to where you register your ontology. Note that it is 'where' not 'if'. To gain the full benefit of your ontology and the ontologies from others, you need to register your creation in a repository.

See the sections below for details on these topics.

a) Implementation considerations

The following should help you define the terms in your new ontology. Remember, you have to make the ontology expandable (scalable) because there will likely be additions. As well, there are a few tricks you can think about before starting.

How to Determine the Terms

To identify the terms of the ontology, you need to first examine the descriptions of your assets, looking for discrete (i.e., non continuous) content. Things that are measured are usually continuous, but the terms that describe these measured things are discrete. For example, alkalinity (or temperature, or salinity) is a continuous measurement, but the term alkalinity is a discrete type of parameter. Another way of telling what might have a discrete description is if you can count the total number of possible descriptions, it is likely to be discrete.

If the possible content of the metadata element is found to be discrete, then it is a likely candidate to be a term in an ontology. For example, if the metadata descriptor was ocean_name and the content was the name of the ocean, then the ocean names could be added to the system as terms in the ontology. In this case, the ontology contains the five ocean names.

Once you have identified those metadata elements that contain discrete terms, you must identify all possible terms that may be contained within those elements. When the term is placed in the element as content, it is considered a value for that element. For the ontology, you should be able to provide a definition of each term, such that its definition is unique to that term. This definition development is a process of building a dictionary of terms for the ontology.

How to create a scalable controlled ontology - allowing for additions

The scalability of an ontology is an important aspect. The ontology should not be limited by the initial terms it contains. To avoid this, you need to examine the terms and think about the general class of things that all the terms are describing. Don't think about an individual term (or an individual car, to extend the vehicle example). Rather, think about the general class of things. Now, attempt to define attributes of the general class. This may not be an easy process. However, if you are successful your ontology will be scalable.

Tips and Tricks - Don't have ontology terms with embedded information

Don’t encode information within the ontology terms. As an example, a term that contains encoded information may have certain characters as meaning certain facts about the term. For example, a single term like XT07aa might indicate an XBT temperature from a T-7 computed using coefficient set aa. Such a term contains information on the type of sensor, the model of sensor, the parameter being measured and processing information. This type of information should be split out of the single term, into multiple terms that apply to different classes of objects in the ontology.

Tips and Tricks - Think about future grouping of terms

At some point, you may realize you have a group of terms that all describe some similar thing. This group of terms may have a single term that describes the group. For example, individual terms of u, v, w may be used to describe the three components of water velocity. As a group, we could describe these three terms as “water current” or “water velocity” or perhaps “ocean currents” (a GCMD term).

Allowing for such groupings to occur in your ontology will help in the management of both the ontology and in the user discovery of the data described by the ontology. Your ontology management should be capable of adding this grouping with minimal impact on the management system.

Tips and Tricks - Don’t allow users to manage the ontology

Users need a mechanism to suggest new terms for the ontology but they cannot be given the ability to add new terms. An ontology is controlled to avoid confusion among terms and to avoid the introduction of errors. Additions, deletions or corrections must be managed by the person responsible for the ontology.

Tips and Tricks - Units are important

The ontology terms used to describe your data may or may not contain explicit units. For example, the terms in the ontology may have a direct association with the unit (i.e., one term can only have one unit). A more preferred method is to allow multiple units for a single term (e.g., distance can have units of meters or kilometers). By allowing multiple units you effectively introduce another type of ontology that your system must support – a unit ontology.

Tips and Tricks - The same syntactic rules

The terms used in the ontology will be created using a set of syntactic rules that may involve capitalization, the use of underscores, or the use of other special characters. The ontology must be developed with consistent application of these rules throughout the ontology terms. Pick your syntactic rules before you start, and stick with them.

Tips and Tricks - Use natural terms

Whenever possible, natural terms that are commonly used within the community, should be used within the ontology. However, if these terms introduce ambiguity, then consider other terms. Unambiguous terms and definitions are the cornerstone of the ontology.

Tips and Tricks - Unambiguous definition

The terms used in your ontology should be associated with rigorous definitions. These definitions should be unambiguous to the community using the ontology.

i. What do people need to think about before creating (or choosing?) and deploying ontologies in their domain?

b. Software Tools for Ontologies
   i. How would someone work with ontologies?
      1. Might be useful to have different starting points up-front, ie “I have a controlled vocabulary” or “I need to choose an ontology” or “I have an ontology but I need to map it to a commonly-used ontology”.
c. Standards and Protocols for Ontologies
   i. This is here because it seems like a good place for the title. Still not sure what exactly needs to go in the section.
d. Ontology Registration and Repositories
   i. Covers the last (and maybe the first?) step of working with ontologies: Registering an ontology if you have one, making use of the repository if you don’t.

Ontolgies Guide Organizing Document

A. Ontologies: Robust Controlled Vocabularies
   1. Introduction: What is an Ontology? [68]
      // paul 11.7.08: Added from III.B. Moved "How ontologies work" from 2.d.
      a. About Ontologies: A Little History
      b. About Ontologies: General Overview
   2. What Ontologies Provide
      a. Document Domain Knowledge [70]
      b. Document and Develop Community Understanding [71]
      c. Organize and Share Data [72]
         // paul 11.7.08: Added content from section III.B.3.
      d. Enable Interoperability [73]
         // paul 11.7.08: Added content from section III.B.2. Needs realignment as interoperability in the old guide covered a lot. Moved "How ontologies work" to 1.
      e. Summary
   3. Core Technologies
      a. Ontology Standards
      b. Storage and Access Fundamentalds
      c. Inference Engines
      d. Software Libraries
   4. Working With Ontologies as a Provider
      // John proposes to write this section
      a. Serving your ontology(ies): Yourself, or through a repository
      b. Creating a Domain Ontology ('Upper Ontology')
      c. Creating Ontologies Using Vocabularies
      d. Registering and Accessing Your Ontology
   5. Working With Ontologies as a User
      // John proposes Paul put this section at the top of his list
      a. Finding them [93]
      b. Citing their terms [94]
      c. Borrowing their terms for your own ontology
      d. Referencing them in your own ontology

Other Resources (not integrated yet):

- Software Tools for Ontologies [636]
- Ontology Tutorials [637]

This site is hosted by the Harte Research Institute, Texas A&M University - Corpus Christi, Corpus Christi, TX, USA
Funding for this web site was provided (long ago) by the National Science Foundation through grant OCE 0607372.
Any material on this site represents the understanding of the authors and contributors to the site, and not necessarily of NSF.

Source URL: https://marinemetadata.org/guides

Links:
[1] https://marinemetadata.org/guides/newuserrecs
[2] mailto:guides@marinemetadata.org
[3] https://marinemetadata.org/guides/guidesabout
[4] http://marinemetadata.org/guides
[5] https://marinemetadata.org/glossary/29#term11033
[6] https://marinemetadata.org/guides/mdataintro/mdatadefined
[7] https://marinemetadata.org/glossary/29#term11037
[8] https://marinemetadata.org/guides/mdataintro/mdatainteroperability
[9] https://marinemetadata.org/glossary/56#term11610
[10] https://marinemetadata.org/guides/mdataintro
[11] https://marinemetadata.org/guides/mdatastandards
[12] https://marinemetadata.org/guides/vocabs
[13] https://marinemetadata.org/glossary/29#term11575
[14] https://marinemetadata.org/guides/vocabs/voctypes/voccat/glossary
[15] https://marinemetadata.org/guides/guidestoc
[16] https://marinemetadata.org/guides
[17] https://marinemetadata.org/guides/glossary
[18] https://marinemetadata.org/guides/acronymglossary
[19] https://marinemetadata.org/guides/mdataintro/mdatadefined/howclassified
[20] https://marinemetadata.org/guides/mdataintro/mdataimportance
[21] https://marinemetadata.org/guides/mdataintro/machinereadability
[22] https://marinemetadata.org/guides/mdataintro/gettingstarted
[23] https://marinemetadata.org/guides/mdataintro/meetingnsfdatareqs
[24] https://marinemetadata.org/guides/mdataintro/writegoodmdata
[25] https://marinemetadata.org/guides/mdataintro/mdataexamples
[26] https://marinemetadata.org/guides/mdatastandards/stdimportance
[27] https://marinemetadata.org/guides/mdatastandards/stdvsspec
[28] https://marinemetadata.org/guides/mdatastandards/comstds
[29] https://marinemetadata.org/guides/mdatastandards/comstds/standardsbodies
[30] https://marinemetadata.org/guides/mdatastandards/comstds/disciplinespecificprofandext
[31] https://marinemetadata.org/guides/mdatastandards/extensionsprofiles
[32] https://marinemetadata.org/guides/mdatastandards/crosswalks
[33] https://marinemetadata.org/guides/mdatastandards/crosswalks/harmonization
[34] https://marinemetadata.org/guides/mdatastandards/crosswalks/semanticmappings
[35] https://marinemetadata.org/guides/mdatastandards/crosswalks/rules
[36] https://marinemetadata.org/guides/mdatastandards/crosswalks/transformation
[37] https://marinemetadata.org/guides/mdatastandards/standardselect
[38] https://marinemetadata.org/guides/vocabs/voctypes
[39] https://marinemetadata.org/guides/vocabs/voctypes/cvusagevsdisc
[40] https://marinemetadata.org/guides/vocabs/voctypes/metadataandcv
[41] https://marinemetadata.org/guides/vocabs/voctypes/cvsemvssyn
[42] https://marinemetadata.org/guides/vocabs/voctypes/voccat
[43] https://marinemetadata.org/guides/vocabs/voctypes/voccat/authorityfile
[44] https://marinemetadata.org/guides/vocabs/voctypes/voccat/dictionary
[45] https://marinemetadata.org/guides/vocabs/voctypes/voccat/gazetteer
[46] https://marinemetadata.org/guides/vocabs/voctypes/voccat/codelist
[47] https://marinemetadata.org/guides/vocabs/voctypes/voccat/taxonomy
[48] https://marinemetadata.org/guides/vocabs/voctypes/voccat/subjectheading
[49] https://marinemetadata.org/guides/vocabs/voctypes/voccat/thesaurus
[50] https://marinemetadata.org/guides/vocabs/voctypes/voccat/semanticnetwork
[51] https://marinemetadata.org/guides/vocabs/voctypes/voccat/ontology
[52] https://marinemetadata.org/guides/vocabs/cvchooseimplement
[53] https://marinemetadata.org/guides/vocabs/cvchooseimplement/cvchoosing
[54] https://marinemetadata.org/guides/vocabs/cvchooseimplement/cvmap
[55] https://marinemetadata.org/guides/vocabs/cvchooseimplement/cvmap/cvmapvine
[56] https://marinemetadata.org/examples/mmihostedwork/ontologieswork/mmivmworkshoptemplate/mmivmworkshoptemplatetoc
[57] https://marinemetadata.org/guides/vocabs/cvchooseimplement/cvsemint
[58] https://marinemetadata.org/guides/vocabs/cvdev
[59] https://marinemetadata.org/guides/vocabs/cvdev/cvdevnew
[60] https://marinemetadata.org/guides/vocabs/cvdev/cvdevnew/cvdevnewterms
[61] https://marinemetadata.org/guides/vocabs/cvdev/cvdevnew/cvdevnewscalable
[62] https://marinemetadata.org/guides/vocabs/cvdev/cvdevnew/cvdevtips
[63] https://marinemetadata.org/guides/vocabs/cvdev/cvdevlegacy
[64] https://marinemetadata.org/guides/vocabs/cvdev/cvdevlegacy/cvdevlegacyapproach
[65] https://marinemetadata.org/guides/vocabs/cvdev/cvdevlegacy/cvdevlegacytips
[66] https://marinemetadata.org/guides/vocabs/cvdev/cvdevlegacy/cvdevlegacyexample
[67] https://marinemetadata.org/guides/vocabs/ont
[68] https://marinemetadata.org/guides/vocabs/ont/definition
[69] https://marinemetadata.org/guides/vocabs/ont/importance
[70] https://marinemetadata.org/guides/vocabs/ont/importance/domainknowledge
[71] https://marinemetadata.org/guides/vocabs/ont/importance/community
[72] https://marinemetadata.org/guides/vocabs/ont/importance/organizeshare
[73] https://marinemetadata.org/guides/vocabs/ont/importance/interoperability
[74] https://marinemetadata.org/guides/vocabs/ont/existing
[75] https://marinemetadata.org/guides/vocabs/ont/existing/finding
[76] https://marinemetadata.org/guides/vocabs/ont/existing/citing
[77] https://marinemetadata.org/guides/vocabs/ont/existing/extending
[78] https://marinemetadata.org/guides/vocabs/ont/provider
[79] https://marinemetadata.org/guides/vocabs/ont/provider/createdomainont
[80] https://marinemetadata.org/guides/vocabs/ont/provider/createvocabont
[81] https://marinemetadata.org/guides/vocabs/ont/provider/servingontologies
[82] https://marinemetadata.org/guides/vocabs/ont/provider/registeraccessont

[83] https://marinemetadata.org/guides/vocabs/ont/coretech
[84] https://marinemetadata.org/guides/vocabs/ont/coretech/standards
[85] https://marinemetadata.org/guides/vocabs/ont/coretech/storage
[86] https://marinemetadata.org/guides/vocabs/ont/coretech/softwarelibs
[87] https://marinemetadata.org/guides/vocabs/ont/coretech/inferencing
[88] https://marinemetadata.org/guides/vocabs/ont/onttut
[89] https://marinemetadata.org/guides/vocabs/vocdef
[90] https://marinemetadata.org/guides/vocabs/vocimportance
[91] https://marinemetadata.org/guides/vocabs/ont/importance/summary
[92] https://marinemetadata.org/guides/vocabs/ont/user
[93] https://marinemetadata.org/guides/vocabs/ont/user/finding
[94] https://marinemetadata.org/guides/vocabs/ont/user/citing
[95] https://marinemetadata.org/guides/vocabs/ont/user/borrow
[96] https://marinemetadata.org/guides/vocabs/ont/user/referencing
[97] https://marinemetadata.org/guides/technicaltools
[98] https://marinemetadata.org/guides/casestudies
[99] https://marinemetadata.org/guides/casestudies/casestudydigarch/view
[100] https://marinemetadata.org/guides/casestudies/casestudycvforharvesting/view
[101] https://marinemetadata.org/guides/casestudies/casestudycv/view
[102] https://marinemetadata.org/guides/otherexternalguides
[103] https://marinemetadata.org/guides/mdataintro/mdatadefined/vocabintro
[104] https://marinemetadata.org/glossary/29#term11039
[105] https://marinemetadata.org/glossary/29#term11055
[106] https://marinemetadata.org/glossary/29#term11574
[107] https://marinemetadata.org/glossary/29#term11044
[108] https://marinemetadata.org/glossary/29#term11046
[109] https://marinemetadata.org/glossary/56#term11615
[110] https://marinemetadata.org/glossary/56#term11644
[111] https://marinemetadata.org/glossary/56#term11596
[112] https://marinemetadata.org/glossary/29#term11032
[113] https://marinemetadata.org/glossary/29#term11027
[114] https://marinemetadata.org/glossary/56#term11586
[115] https://marinemetadata.org/glossary/29#term11038
[116] https://marinemetadata.org/guides/mdatastandards/extensionsprofiles#profile
[117] https://marinemetadata.org/glossary/29#term11034
[118] https://marinemetadata.org/guides/mdataintro/mdatadefined#element
[119] https://marinemetadata.org/glossary/29#term11041
[120] https://marinemetadata.org/guides/mdataintro/mdatadefined#value
[121] https://marinemetadata.org/glossary/56#term11634
[122] https://marinemetadata.org/glossary/29#term11024
[123] https://marinemetadata.org/glossary/29#term11057
[124] https://marinemetadata.org/glossary/29#term11026
[125] https://marinemetadata.org/glossary/29#term11047
[126] https://marinemetadata.org/references/nisometa
[127] https://marinemetadata.org/glossary/56#term11621
[128] https://marinemetadata.org/glossary/56#term11604
[129] https://marinemetadata.org/glossary/56#term11598
[130] https://marinemetadata.org/glossary/29#term11058
[131] https://marinemetadata.org/glossary/56#term11649
[132] https://marinemetadata.org/glossary/29#term11045
[133] http://office.microsoft.com/en-us/excel/
[134] http://freemind.sourceforge.net/wiki/index.php/Main_Page
[135] https://marinemetadata.org/glossary/56#term11706
[136] https://marinemetadata.org/56#term11582
[137] https://marinemetadata.org/glossary/29#term719
[138] http://www.nsf.gov/
[139] http://www.nsf.gov/pubs/2004/nsf04004/print_toc.htm
[140] https://marinemetadata.org/news/nsfdataplan
[141] http://www.nsf.gov/pubs/2004/nsf04004/nsf04004_3.htm
[142] http://www.nodc.noaa.gov/
[143] http://www.nodc.noaa.gov/General/NODC-Submit/
[144] http://www.nodc.noaa.gov/General/NODC-Submit/submit-guide.html
[145] http://lwf.ncdc.noaa.gov/oa/ncdc.html
[146] http://www.ncdc.noaa.gov/paleo/contrib.html
[147] http://www.ncdc.noaa.gov/paleo/text/tringguide.html
[148] http://www.ncdc.noaa.gov/paleo/impd/impd_trscr_contrib.html
[149] http://www.ngdc.noaa.gov/ngdc.html
[150] http://www.ngdc.noaa.gov/mgg/bathymetry/iho.html
[151] http://nsidc.org/
[152] http://nsidc.org/data/submit.html
[153] http://cdiac.esd.ornl.gov/home.html
[154] http://cdiac.ornl.gov/datasubmission.html
[155] http://daac.ornl.gov/PI/bestprac.html
[156] http://cdiac.ornl.gov/staff.html
[157] https://marinemetadata.org/glossary/29#term11040
[158] https://marinemetadata.org/guides/metadatastandards
[159] https://marinemetadata.org/guides/otherexternalguides/cscguide
[160] https://marinemetadata.org/glossary/29#term11036
[161] http://sofia.usgs.gov/metadata/sflwww/ever_hydro_wq_data.xml
[162] https://marinemetadata.org/glossary/56#term11601
[163] http://sofia.usgs.gov/metadata/sflwww/sofia_website_04.html
[164] http://sofia.usgs.gov/metadata/sflwww/ever_hydro_wq_data.html
[165] http://estuarinebathymetry.noaa.gov/documentation/30m_bathy.html
[166] http://www.mbari.org/muse/metadata/Processed/sst.html
[167] http://www.mbari.org/muse/metadata/Processed/nighttrack.html
[168] http://www.csc.noaa.gov/crs/real_time/hab/sst_gomx_1.html
[169] http://www.chesapeakebay.net/dataandtools.aspx
[170] http://gcmd.gsfc.nasa.gov/KeywordSearch/Metadata.do?Portal=GCMD_Services&
amp;KeywordPath=ServiceParameters%7CMETADATA+HANDLING%7CMETADATA+TRANSFORMATION%2FCONVERSION&EntryId=Glob_Chng_Mast_Direct&
amp;MetadataView=Brief&MetadataType=1&lbnode=gcmd3

[171] http://pubs.usgs.gov/of/2000/of00-304/lisound/data/chap09/buzas.htm
[172] http://walrus.wr.usgs.gov/infobank/b/b103pr/html/b-1-03-pr.meta.html
[173] http://www.mersea.eu.org/html/information/catalog/products/CORIOLIS-GLOBAL-NRT-OBS.html
[174] https://marinemetadata.org/conventions/content-standards
[175] https://marinemetadata.org/glossary/29#term11029
[176] https://marinemetadata.org/glossary/29#term11572
[177] https://marinemetadata.org/guides/mdatastandards#content
[178] https://marinemetadata.org/glossary/29#term11573
[179] https://marinemetadata.org/guides/mdatastandards#format
[180] https://marinemetadata.org/glossary/29#term11747
[181] https://marinemetadata.org/glossary/56#term11614
[182] https://marinemetadata.org/glossary/56#term11585
[183] https://marinemetadata.org/glossary/56#term11707
[184] https://marinemetadata.org/glossary/56#term11708
[185] https://marinemetadata.org/glossary/29#term11576
[186] https://marinemetadata.org/glossary/29#term11577
[187] https://marinemetadata.org/glossary/56#term11627
[188] https://marinemetadata.org/glossary/56#term11590
[189] https://marinemetadata.org/guides/mdataintro/mdatadefined/howclassified#discovery
[190] https://marinemetadata.org/glossary/56#term11588
[191] http://www.iso.org/
[192] http://www.iso.ch/iso/en/aboutiso/isomembers/index.html
[193] http://www.ansi.org/
[194] http://www.niso.org/standards/standard_detail.cfm?std_id=725
[195] http://www.fgdc.gov/metadata
[196] https://marinemetadata.org/glossary/29#term11571
[197] http://marinemetadata.org/references/fgdcnbii
[198] http://marinemetadata.org/references/wmometadata
[199] http://marinemetadata.org/references/marineprofile19115
[200] http://marinemetadata.org/guides/mdatastandards/crosswalks/harmonization
[201] http://marinemetadata.org/guides/mdatastandards/crosswalks/semanticmappings
[202] http://marinemetadata.org/guides/mdatastandards/crosswalks/rules
[203] http://marinemetadata.org/guides/mdatastandards/crosswalks/transformation
[204] http://marinemetadata.org/taxonomy/term/11681
[205] http://marinemetadata.org/taxonomy/term/11677
[206] http://marinemetadata.org/taxonomy/term/11657
[207] http://marinemetadata.org/guides/mdatastandards/crosswalks
[208] https://marinemetadata.org/glossary/29#term11030
[209] https://marinemetadata.org/glossary/29#term11748
[210] https://marinemetadata.org/glossary/56#term11631
[211] http://www.w3.org/2004/02/skos/mapping/spec/
[212] https://marinemetadata.org/glossary/56#term11639
[213] http://marinemetadata.org/node/6346
[214] https://marinemetadata.org/glossary/29#term11053
[215] https://marinemetadata.org/glossary/29#term11054
[216] http://www.w3.org/2004/02/skos/mapping/spec/2004-11-11.html
[217] https://marinemetadata.org/glossary/29#term11059
[218] https://marinemetadata.org/glossary/29#term11056
[219] https://marinemetadata.org/glossary/29#term11025
[220] https://marinemetadata.org/glossary/56#term11662
[221] https://marinemetadata.org/glossary/29#term11060
[222] http://www.clir.org/pubs/reports/pub91/1knowledge.html
[223] https://marinemetadata.org/glossary/29#term11570
[224] https://marinemetadata.org/glossary/29#term11578
[225] https://marinemetadata.org/glossary/29#term11023
[226] https://marinemetadata.org/glossary/29#term11579
[227] https://marinemetadata.org/glossary/56#term11635
[228] http://marinemetadata.org/glossary/29#letterm
[229] http://marinemetadata.org/glossary/29#letters
[230] http://marinemetadata.org/glossary/29#letterc
[231] http://marinemetadata.org/glossary/29
[232] https://marinemetadata.org/glossary/56#term11582
[233] https://marinemetadata.org/glossary/56#term11580
[234] https://marinemetadata.org/glossary/29#term11028
[235] https://marinemetadata.org/guides/vocabs/voctypes/voccat#flat
[236] https://marinemetadata.org/glossary/29#term11048
[237] https://marinemetadata.org/guides/vocabs/voctypes/voccat#relational
[238] https://marinemetadata.org/node/2084
[239] https://marinemetadata.org/node/2082
[240] https://marinemetadata.org/node/2083
[241] https://marinemetadata.org/node/2081
[242] https://marinemetadata.org/node/2088
[243] https://marinemetadata.org/node/2087
[244] https://marinemetadata.org/node/2089
[245] https://marinemetadata.org/node/2086
[246] https://marinemetadata.org/node/2085
[247] http://marinemetadata.org/references/dublincore
[248] https://marinemetadata.org/node/2080
[249] https://marinemetadata.org/references/siogdcterms
[250] https://marinemetadata.org/references/cdivocabs
[251] https://marinemetadata.org/glossary/29#term11042
[252] https://marinemetadata.org/guides/vocabs/voctypes/voccat/#multilevel
[253] http://earth-info.nga.mil/gns/html/cntry_files.html
[254] https://marinemetadata.org/references/gcmdsensors
[255] http://www.marinemetadata.org/node/2080
[256] http://www.marinemetadata.org/node/2084
[257] http://www.marinemetadata.org/node/2082
[258] http://www.marinemetadata.org/node/2083
[259] http://www.marinemetadata.org/node/2081
[260] http://www.marinemetadata.org/node/2087

[261] http://www.marinemetadata.org/node/2089
[262] http://www.marinemetadata.org/node/2086
[263] http://www.marinemetadata.org/node/2085
[264] http://www.techstreet.com/standards/NISO/Z39_19_2003?product_id=1171385
[265] http://alexandria.ucsb.edu/gazetteer/
[266] http://www.usgs.gov/science/about/term-simple.php
[267] http://www.usgs.gov/science/about/faq/
[268] https://marinemetadata.org/guides/vocabs/cvchooseimplement/cvimp
[269] https://marinemetadata.org/glossary/29#term11051
[270] https://marinemetadata.org/semanticframework
[271] https://marinemetadata.org/guides/vocabs/cvchooseimplement/cvsemint#framework
[272] http://gcmd.nasa.gov/User/difguide/parameters.html
[273] http://sweet.jpl.nasa.gov/ontology/
[274] http://www.iode.org/index.php?Itemid=65&id=24&option=com_content&task=view/
[275] http://gcmd.nasa.gov/
[276] http://www.seadatanet.org/
[277] http://www.catalogueoflife.org/annual-checklist/info_source_dbs.php
[278] mailto:ask@marinemetadata.org
[279] https://marinemetadata.org/glossary/56#term11640
[280] http://marinemetadata.org/references/voc2owl
[281] http://mmisw.org/voc2rdf
[282] http://marinemetadata.org/references/vine
[283] https://marinemetadata.org/tools
[284] https://marinemetadata.org/workshoptemp
[285] http://mmisw.org/or
[286] https://marinemetadata.org/glossary/29#term11035
[287] https://marinemetadata.org/guides/mdatastandards/extensionsprofiles#extension
[288] https://marinemetadata.org/glossary/56#term11622
[289] https://marinemetadata.org/node/4944
[290] https://marinemetadata.org/references/vine
[291] https://marinemetadata.org/node/4962
[292] https://marinemetadata.org/glossary/29#term11052
[293] https://marinemetadata.org/references/coardscf
[294] http://marinemetadata.org/guides/vocabs/cvchooseimplement
[295] http://marinemetadata.org/guides/vocabs/cvdev
[296] https://marinemetadata.org/glossary/29#term11050
[297] https://marinemetadata.org/glossary/56#term11648
[298] https://marinemetadata.org/guides/dictionary#vocabulary
[299] https://marinemetadata.org/ask
[300] http://marinemetadata.org/glossary/29#letterp
[301] http://marinemetadata.org/glossary/29#letterv
[302] http://en.wikipedia.org/wiki/Marsden_square
[303] http://marinemetadata.org/glossary/29#term11041
[304] http://marinemetadata.org/glossary/29#term11057
[305] https://marinemetadata.org/glossary/29#term11557
[306] https://marinemetadata.org/glossary/56#term11638
[307] http://openioos.org/
[308] http://www.opengeospatial.org/projects/groups/sensorweb
[309] http://www.orionprogram.org/OOI/default.html
[310] http://www.oceanobservatories.org/
[311] http://marinemetadata.org/node/2248
[312] https://marinemetadata.org/glossary/56#term11645
[313] https://marinemetadata.org/glossary/56#term11608
[314] https://marinemetadata.org/glossary/56#term11630
[315] https://marinemetadata.org/glossary/56#term11583
[316] http://marinemetadata.org/references/sensorsontmmi
[317] https://marinemetadata.org/references/rdf
[318] https://marinemetadata.org/references/owl
[319] http://mmisw.org/
[320] http://ican.science.oregonstate.edu/
[321] https://marinemetadata.org/examples/mmihostedwork/demos/ican
[322] https://marinemetadata.org/glossary/56#term11597
[323] http://mmisw.org/orr/
[324] http://bioportal.bioontology.org/
[325] http://bioportal.nci.nih.gov/ncbo/faces/index.xhtml
[326] http://olp.dfki.de/ontoselect/
[327] http://www.daml.org/ontologies/
[328] http://protegewiki.stanford.edu/index.php/Protege_Ontology_Library
[329] http://swoogle.umbc.edu/
[330] http://www.ontosearch.org/
[331] http://iws.seu.edu.cn/services/falcons/objectsearch/index.jsp
[332] http://watson.kmi.open.ac.uk/WatsonWUI/
[333] http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/SemanticWebSearchEngines
[334] http://esw.w3.org/topic/SemanticWebTools#head-09ab45ef1fb5d28af30c2787c5ef6705329abc5d
[335] http://marinemetadata.org/apguides/ontprovidersguide/ontguideconstructinguris
[336] http://mmisw.org/ont/mmi/20081116T071659/sensor
[337] http://mmisw.org/ont/mmi/20081116T071659/sensor/ProcessOutput
[338] http://aabbcc112233.org
[339] http://ontologiesforscience.org
[340] http://www.w3.org/TR/owl-guide/#Namespaces
[341] http://www.w3.org/TR/xml-names/
[342] https://marinemetadata.org/node/6673
[343] https://marinemetadata.org/node/6674
[344] https://marinemetadata.org/node/6672
[345] https://marinemetadata.org/node/6675
[346] http://marinemetadata.org/references/protege
[347] http://www.topquadrant.com/products/TB_Composer.html
[348] http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html
[349] http://www.jfsowa.com/ontology/guided.htm
[350] http://www.jfsowa.com/ontology/

[351] http://www.obitko.com/tutorials/ontologies-semantic-web/
[352] http://mmisw.org/mmi/examples/sea_surface_temperature
[353] https://marinemetadata.org/references/protege
[354] https://marinemetadata.org/workshoptemp/pre/tools
[355] http://mywebsite.com/ontologies/
[356] https://marinemetadata.org/glossary/56#term11641
[357] https://marinemetadata.org/workshoptemp/pre/owlont
[358] https://marinemetadata.org/apguides/ontprovidersguide
[359] https://marinemetadata.org/apguides/ontprovidersguide/ontguideconstructinguris
[360] http://www.w3.org/TR/swbp-vocab-pub/
[361] http://www.w3.org/
[362] http://www.w3.org/Consortium/
[363] http://vast.uah.edu/SensorML
[364] http://schemas.opengis.net/sensorML/1.0.1/base.xsd
[365] http://www.w3.org/XML/
[366] http://www.w3schools.com/xml/default.asp
[367] http://www.w3.org/RDF/
[368] http://planetrdf.com/guide/
[369] https://marinemetadata.org/glossary/56#term11589
[370] http://www.w3.org/TR/owl-ref/
[371] http://www.w3.org/TR/owl-guide/
[372] http://protege.stanford.edu/plugins/owl/api
[373] http://jena.sourceforge.net/
[374] http://semanticweb.org/wiki/RDFReactor
[375] http://www.talis.com/platform
[376] http://esw.w3.org/topic/SemanticWebTools#head-242527e5721cb193f7a9d13af414deb6dc213c9a
[377] http://marinemetadata.org/guides/vocabs/ont/coretech/softwarelibs
[378] http://esw.w3.org/topic/SemanticWebTools#head-d07454b4f0d51f5e9d878822d911d0bfea9dcdfd
[379] https://marinemetadata.org/examples/mmihostedwork/ontologieswork/tutorials/ontologycreationintroguides
[380] https://marinemetadata.org/community/teams/ont/tutorials/tutorprotegeowl
[381] https://marinemetadata.org/examples/mmihostedwork/ontologieswork/ontwebservices/ontwebservicestutorial
[382] https://marinemetadata.org/examples/mmihostedwork/ontologieswork/guidesontologyprovider/guideontprovider20070205/view
[383] https://marinemetadata.org/toolsDIVE
[384] http://www.celiktabela.com
[385] https://marinemetadata.org/files/mmi/CaseStudy_DIGARCH_Cruise_Harvest.pdf
[386] https://marinemetadata.org/files/mmi/CaseStudy_CV_for_metadata_harvest.pdf
[387] https://marinemetadata.org/files/mmi/CaseStudy_CVdictionary_20070605.pdf
[388] https://marinemetadata.org/system/files/CaseStudy_CV_for_metadata_harvest.pdf
[389] https://marinemetadata.org/system/files/CaseStudy_CVdictionary_20070605.pdf
[390] https://marinemetadata.org/system/files/CaseStudy_DIGARCH_Cruise_Harvest.pdf
[391] https://marinemetadata.org/node/add/mmiref
[392] https://marinemetadata.org/references/ecolont
[393] https://marinemetadata.org/taxonomy/term/129
[394] https://marinemetadata.org/taxonomy/term/139
[395] https://marinemetadata.org/taxonomy/term/190
[396] https://marinemetadata.org/taxonomy/term/170
[397] https://marinemetadata.org/references/geoncisummerinstitut
[398] https://marinemetadata.org/taxonomy/term/11665
[399] https://marinemetadata.org/taxonomy/term/11666
[400] https://marinemetadata.org/taxonomy/term/11668
[401] https://marinemetadata.org/taxonomy/term/11669
[402] https://marinemetadata.org/references/gettylibraryprimer
[403] https://marinemetadata.org/references/datamgtcookbookgsdi
[404] https://marinemetadata.org/references/sgxml
[405] https://marinemetadata.org/taxonomy/term/11667
[406] https://marinemetadata.org/references/cscguide
[407] https://marinemetadata.org/references/nsdlprimer
[408] https://marinemetadata.org/references/nsgicprimer
[409] https://marinemetadata.org/taxonomy/term/144
[410] https://marinemetadata.org/taxonomy/term/130
[411] https://marinemetadata.org/references/metadata-demystified
[412] https://marinemetadata.org/references/scholarlycitation
[413] https://marinemetadata.org/references/ontology101
[414] https://marinemetadata.org/guides/otherexternalguides/oceanteacherlink
[415] https://marinemetadata.org/references/datacitationvdc
[416] https://marinemetadata.org/references/w3cbpsemweb
[417] https://marinemetadata.org/references/w3curiurnurl
[418] https://marinemetadata.org/admin/content/taxonomy/29/add/term
[419] https://marinemetadata.org/admin/content/taxonomy/edit/vocabulary/29?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[420] https://marinemetadata.org/admin/settings/glossary?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[421] https://marinemetadata.org/glossary/29#lettera
[422] https://marinemetadata.org/glossary/29#letterc
[423] https://marinemetadata.org/glossary/29#letterd
[424] https://marinemetadata.org/glossary/29#letterf
[425] https://marinemetadata.org/glossary/29#letterg
[426] https://marinemetadata.org/glossary/29#letterh
[427] https://marinemetadata.org/glossary/29#letteri
[428] https://marinemetadata.org/glossary/29#letterm
[429] https://marinemetadata.org/glossary/29#lettero
[430] https://marinemetadata.org/glossary/29#letterp
[431] https://marinemetadata.org/glossary/29#letterr
[432] https://marinemetadata.org/glossary/29#letters
[433] https://marinemetadata.org/glossary/29#lettert
[434] https://marinemetadata.org/glossary/29#letteru
[435] https://marinemetadata.org/glossary/29#letterv
[436] https://marinemetadata.org/glossary/29#letterw
[437] https://marinemetadata.org/admin/content/taxonomy/edit/term/11570?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[438] https://marinemetadata.org/admin/content/taxonomy/edit/term/719?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[439] https://marinemetadata.org/admin/content/taxonomy/edit/term/11571?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[440] https://marinemetadata.org/admin/content/taxonomy/edit/term/11572?destination=print%2Fbook%2Fexport%2Fhtml%2F2093

[441] https://marinemetadata.org/admin/content/taxonomy/edit/term/11024?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[442] https://marinemetadata.org/admin/content/taxonomy/edit/term/11025?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[443] https://marinemetadata.org/admin/content/taxonomy/edit/term/11049?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[444] https://marinemetadata.org/admin/content/taxonomy/edit/term/11026?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[445] https://marinemetadata.org/admin/content/taxonomy/edit/term/11027?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[446] https://marinemetadata.org/admin/content/taxonomy/edit/term/11028?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[447] https://marinemetadata.org/admin/content/taxonomy/edit/term/11029?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[448] https://marinemetadata.org/admin/content/taxonomy/edit/term/11573?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[449] https://marinemetadata.org/admin/content/taxonomy/edit/term/11574?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[450] https://marinemetadata.org/admin/content/taxonomy/edit/term/11575?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[451] https://marinemetadata.org/admin/content/taxonomy/edit/term/11030?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[452] https://marinemetadata.org/admin/content/taxonomy/edit/term/11031?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[453] https://marinemetadata.org/admin/content/taxonomy/edit/term/11032?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[454] https://marinemetadata.org/admin/content/taxonomy/edit/term/11033?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[455] https://marinemetadata.org/admin/content/taxonomy/edit/term/11023?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[456] https://marinemetadata.org/admin/content/taxonomy/edit/term/11034?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[457] https://marinemetadata.org/admin/content/taxonomy/edit/term/11035?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[458] https://marinemetadata.org/admin/content/taxonomy/edit/term/11036?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[459] https://marinemetadata.org/admin/content/taxonomy/edit/term/11037?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[460] https://marinemetadata.org/admin/content/taxonomy/edit/term/11747?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[461] https://marinemetadata.org/admin/content/taxonomy/edit/term/11038?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[462] https://marinemetadata.org/admin/content/taxonomy/edit/term/11039?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[463] https://marinemetadata.org/admin/content/taxonomy/edit/term/11040?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[464] https://marinemetadata.org/admin/content/taxonomy/edit/term/11041?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[465] https://marinemetadata.org/admin/content/taxonomy/edit/term/11042?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[466] https://marinemetadata.org/admin/content/taxonomy/edit/term/11043?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[467] https://marinemetadata.org/admin/content/taxonomy/edit/term/11044?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[468] https://marinemetadata.org/admin/content/taxonomy/edit/term/11045?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[469] https://marinemetadata.org/admin/content/taxonomy/edit/term/11046?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[470] https://marinemetadata.org/admin/content/taxonomy/edit/term/11048?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[471] https://marinemetadata.org/admin/content/taxonomy/edit/term/11047?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[472] https://marinemetadata.org/admin/content/taxonomy/edit/term/11050?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[473] https://marinemetadata.org/admin/content/taxonomy/edit/term/11576?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[474] https://marinemetadata.org/admin/content/taxonomy/edit/term/11051?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[475] https://marinemetadata.org/admin/content/taxonomy/edit/term/11052?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[476] https://marinemetadata.org/admin/content/taxonomy/edit/term/11053?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[477] https://marinemetadata.org/admin/content/taxonomy/edit/term/11579?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[478] https://marinemetadata.org/admin/content/taxonomy/edit/term/11557?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[479] https://marinemetadata.org/admin/content/taxonomy/edit/term/11054?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[480] https://marinemetadata.org/admin/content/taxonomy/edit/term/11577?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[481] https://marinemetadata.org/admin/content/taxonomy/edit/term/11578?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[482] https://marinemetadata.org/admin/content/taxonomy/edit/term/11055?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[483] https://marinemetadata.org/admin/content/taxonomy/edit/term/11056?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[484] https://marinemetadata.org/admin/content/taxonomy/edit/term/11057?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[485] https://marinemetadata.org/admin/content/taxonomy/edit/term/11058?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[486] https://marinemetadata.org/admin/content/taxonomy/edit/term/11059?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[487] https://marinemetadata.org/admin/content/taxonomy/edit/term/11748?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[488] https://marinemetadata.org/admin/content/taxonomy/edit/term/11060?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[489] http://dublincore.org/documents/usageguide/glossary.shtml
[490] http://www.getty.edu/research/conducting_research/standards/intrometadata/glossary.html
[491] https://marinemetadata.org/admin/content/taxonomy/56/add/term
[492] https://marinemetadata.org/admin/content/taxonomy/edit/vocabulary/56?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[493] https://marinemetadata.org/glossary/56#lettera
[494] https://marinemetadata.org/glossary/56#letterb
[495] https://marinemetadata.org/glossary/56#letterc
[496] https://marinemetadata.org/glossary/56#letterd
[497] https://marinemetadata.org/glossary/56#lettere
[498] https://marinemetadata.org/glossary/56#letterf
[499] https://marinemetadata.org/glossary/56#letterg
[500] https://marinemetadata.org/glossary/56#letterh
[501] https://marinemetadata.org/glossary/56#letteri
[502] https://marinemetadata.org/glossary/56#letterl
[503] https://marinemetadata.org/glossary/56#letterm
[504] https://marinemetadata.org/glossary/56#lettern
[505] https://marinemetadata.org/glossary/56#lettero
[506] https://marinemetadata.org/glossary/56#letterp
[507] https://marinemetadata.org/glossary/56#letterq
[508] https://marinemetadata.org/glossary/56#letterr
[509] https://marinemetadata.org/glossary/56#letters
[510] https://marinemetadata.org/glossary/56#letteru
[511] https://marinemetadata.org/glossary/56#letterv
[512] https://marinemetadata.org/glossary/56#letterw
[513] https://marinemetadata.org/glossary/56#letterx
[514] https://marinemetadata.org/admin/content/taxonomy/edit/term/11708?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[515] https://marinemetadata.org/admin/content/taxonomy/edit/term/11580?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[516] https://marinemetadata.org/admin/content/taxonomy/edit/term/11581?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[517] https://marinemetadata.org/admin/content/taxonomy/edit/term/11582?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[518] https://marinemetadata.org/admin/content/taxonomy/edit/term/11583?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[519] https://marinemetadata.org/admin/content/taxonomy/edit/term/11584?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[520] https://marinemetadata.org/admin/content/taxonomy/edit/term/11585?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[521] https://marinemetadata.org/admin/content/taxonomy/edit/term/11706?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[522] https://marinemetadata.org/admin/content/taxonomy/edit/term/11586?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[523] https://marinemetadata.org/admin/content/taxonomy/edit/term/11587?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[524] https://marinemetadata.org/admin/content/taxonomy/edit/term/11588?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[525] https://marinemetadata.org/admin/content/taxonomy/edit/term/11707?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[526] https://marinemetadata.org/admin/content/taxonomy/edit/term/11589?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[527] https://marinemetadata.org/admin/content/taxonomy/edit/term/11590?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[528] https://marinemetadata.org/admin/content/taxonomy/edit/term/11591?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[529] https://marinemetadata.org/admin/content/taxonomy/edit/term/11592?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[530] https://marinemetadata.org/admin/content/taxonomy/edit/term/11593?destination=print%2Fbook%2Fexport%2Fhtml%2F2093

[531] https://marinemetadata.org/admin/content/taxonomy/edit/term/11594?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[532] https://marinemetadata.org/admin/content/taxonomy/edit/term/11595?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[533] https://marinemetadata.org/admin/content/taxonomy/edit/term/11596?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[534] https://marinemetadata.org/admin/content/taxonomy/edit/term/11597?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[535] https://marinemetadata.org/admin/content/taxonomy/edit/term/11598?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[536] https://marinemetadata.org/admin/content/taxonomy/edit/term/11599?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[537] https://marinemetadata.org/admin/content/taxonomy/edit/term/11600?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[538] https://marinemetadata.org/admin/content/taxonomy/edit/term/11601?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[539] https://marinemetadata.org/admin/content/taxonomy/edit/term/11602?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[540] https://marinemetadata.org/admin/content/taxonomy/edit/term/11603?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[541] https://marinemetadata.org/admin/content/taxonomy/edit/term/11604?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[542] https://marinemetadata.org/admin/content/taxonomy/edit/term/11605?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[543] https://marinemetadata.org/admin/content/taxonomy/edit/term/11606?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[544] https://marinemetadata.org/admin/content/taxonomy/edit/term/11607?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[545] https://marinemetadata.org/admin/content/taxonomy/edit/term/11608?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[546] https://marinemetadata.org/admin/content/taxonomy/edit/term/11609?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[547] https://marinemetadata.org/admin/content/taxonomy/edit/term/11610?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[548] https://marinemetadata.org/admin/content/taxonomy/edit/term/11611?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[549] https://marinemetadata.org/admin/content/taxonomy/edit/term/11612?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[550] https://marinemetadata.org/admin/content/taxonomy/edit/term/11613?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[551] https://marinemetadata.org/admin/content/taxonomy/edit/term/11614?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[552] https://marinemetadata.org/admin/content/taxonomy/edit/term/11615?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[553] https://marinemetadata.org/admin/content/taxonomy/edit/term/11616?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[554] https://marinemetadata.org/admin/content/taxonomy/edit/term/11617?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[555] https://marinemetadata.org/admin/content/taxonomy/edit/term/11662?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[556] https://marinemetadata.org/admin/content/taxonomy/edit/term/11618?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[557] https://marinemetadata.org/admin/content/taxonomy/edit/term/11619?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[558] https://marinemetadata.org/admin/content/taxonomy/edit/term/11620?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[559] https://marinemetadata.org/admin/content/taxonomy/edit/term/11621?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[560] https://marinemetadata.org/admin/content/taxonomy/edit/term/11622?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[561] https://marinemetadata.org/admin/content/taxonomy/edit/term/11623?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[562] https://marinemetadata.org/admin/content/taxonomy/edit/term/11624?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[563] https://marinemetadata.org/admin/content/taxonomy/edit/term/11625?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[564] https://marinemetadata.org/admin/content/taxonomy/edit/term/11626?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[565] https://marinemetadata.org/admin/content/taxonomy/edit/term/11627?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[566] https://marinemetadata.org/admin/content/taxonomy/edit/term/11628?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[567] https://marinemetadata.org/admin/content/taxonomy/edit/term/11629?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[568] https://marinemetadata.org/admin/content/taxonomy/edit/term/11630?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[569] https://marinemetadata.org/admin/content/taxonomy/edit/term/11631?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[570] https://marinemetadata.org/admin/content/taxonomy/edit/term/11632?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[571] https://marinemetadata.org/admin/content/taxonomy/edit/term/11633?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[572] https://marinemetadata.org/admin/content/taxonomy/edit/term/11634?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[573] https://marinemetadata.org/admin/content/taxonomy/edit/term/11635?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[574] https://marinemetadata.org/admin/content/taxonomy/edit/term/11636?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[575] https://marinemetadata.org/admin/content/taxonomy/edit/term/11637?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[576] https://marinemetadata.org/admin/content/taxonomy/edit/term/11638?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[577] https://marinemetadata.org/admin/content/taxonomy/edit/term/11639?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[578] https://marinemetadata.org/admin/content/taxonomy/edit/term/11640?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[579] https://marinemetadata.org/admin/content/taxonomy/edit/term/11641?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[580] https://marinemetadata.org/admin/content/taxonomy/edit/term/11642?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[581] https://marinemetadata.org/admin/content/taxonomy/edit/term/11643?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[582] https://marinemetadata.org/admin/content/taxonomy/edit/term/11644?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[583] https://marinemetadata.org/admin/content/taxonomy/edit/term/11646?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[584] https://marinemetadata.org/admin/content/taxonomy/edit/term/11645?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[585] https://marinemetadata.org/admin/content/taxonomy/edit/term/11647?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[586] https://marinemetadata.org/admin/content/taxonomy/edit/term/11648?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[587] https://marinemetadata.org/admin/content/taxonomy/edit/term/11649?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[588] https://marinemetadata.org/admin/content/taxonomy/edit/term/11650?destination=print%2Fbook%2Fexport%2Fhtml%2F2093
[589] http://www.whoi.edu/
[590] http://www.mbari.org/
[591] http://www.drdc-rddc.gc.ca/index1-eng.asp
[592] http://www.usgs.gov/
[593] http://www-sio.ucsd.edu/
[594] http://www.sdsc.edu/
[595] https://marinemetadata.org/guides/dictionary#controlledvocabulary
[596] https://marinemetadata.org/guides/dictionary#dictionary
[597] https://marinemetadata.org/guides/dictionary#thesaurus
[598] https://marinemetadata.org/guides/dictionary#class
[599] https://marinemetadata.org/guides/dictionary#relationship
[600] https://marinemetadata.org/examples/mmihostedwork/ontologieswork/tutorials/tutorprotegeowl
[601] https://marinemetadata.org/guides/dictionary#semanticinteroperability
[602] http://sourceforge.net/project/showfiles.php?group_id=170123
[603] https://marinemetadata.org/images/naturalcatastrophe
[604] http://lisi.insa-lyon.fr/%7Elaurini
[605] https://marinemetadata.org/transport/refsall/rdf
[606] http://mmisw.org/semor/
[607] https://marinemetadata.org/node/2248
[608] https://marinemetadata.org/references/sensorsontmmi
[609] https://marinemetadata.org/references/cmaptools
[610] http://www.mindswap.org/downloads/
[611] http://protege.stanford.edu/
[612] http://www.cs.man.ac.uk/%7Ehorrocks/FaCT/
[613] http://pellet.owldl.com/
[614] http://www.topquadrant.com/topbraid/composer/index.html
[615] http://www.w3.org/2001/11/IsaViz/
[616] http://codip.grci.com/Tools/Components.html
[617] http://www.daml.org/
[618] http://www.ontoknowledge.org/oil/
[619] https://marinemetadata.org/references/voc2owl
[620] http://www.cs.yale.edu/homes/dvm/daml/ontology-translation.html

[621] http://www.geneontology.org/GO.tools.shtml
[622] http://www.xml.com/pub/a/2004/07/14/onto.html
[623] https://marinemetadata.org/tools/find/2922
[624] https://marinemetadata.org/info/aboutmmi/planning/website/techteam/diveteam/divetools
[625] http://marinemetadata.org/community/teams/guides/generaldocuments/howtoaddguide#new
[626] http://ontolog.cim3.net/cgi-bin/wiki.pl?SICoP
[627] https://marinemetadata.org/node/5798
[628] https://marinemetadata.org/semanticframeworkconcept/sfconceptofoperations
[629] http://www.w3.org/TR/2003/PR-owl-guide-20031209/wine
[630] http://www.ksl.stanford.edu/people/dlm/webont/wineAgent/
[631] http://example.org/
[632] http://en.wikipedia.org/wiki/Uniform_Resource_Identifier#Relationship_to_URL_and_URN
[633] http://www.w3.org/TR/uri-clarification/
[634] https://marinemetadata.org/files/mmi/images/wine_asserted.png
[635] https://marinemetadata.org/files/mmi/images/wine_inferred.png
[636] https://marinemetadata.org/guides/vocabs/ontologies/ontsoftware
[637] https://marinemetadata.org/guides/vocabs/ontologies/onttut