

	<b>&lt;MMI /&gt; Case Study</b>
<b>Title</b>	<b>Controlled Vocabulary Dictionary (cvdict) for ships and submergence vehicles</b>
<b>Website</b>	<a href="http://gdccoll.ucsd.edu:8080/digarch">http://gdccoll.ucsd.edu:8080/digarch</a>
<b>Contact</b>	Stephen Miller
<b>Organization</b>	Scripps Institution of Oceanography
<b>Audience</b>	Digital data collection managers and developers; supporting queries by researchers and students
<b>Discipline</b>	Marine Geology and Geophysics, also supports Physical, Biological and Chemical Oceanography, Acoustics; approach is relevant to any discipline
<b>Problem</b>	<p>Across disciplines, and over decades, one of the most troublesome obstacles to wide adoption of metadata has been the lack of basic <b>definitions</b> for all approved metadata values. The lack of controlled vocabularies with definitions is a barrier to interoperability.</p> <p>In many projects today definitions for standard parameter names may exist, and lists of approved values may exist, but it is rare to find definitions for all the approved metadata values. As a consequence, collection builders or users at other institutions or in other disciplines attempting to make use of the metadata or data frequently need to launch emails or phone calls to find out what is really meant, and considerable time and effort is wasted.</p>
<b>Solution</b>	A practical solution involves the publishing of a Controlled Vocabulary Dictionary (cvdict), with a record for every authorized metadata value in the project. Each record has three fields, in the syntax: <formal parameter name>,"<authorized value>","brief definition of the authorized value>"
<b>Input resources</b>	<p>1. The definitions for each formal metadata parameter name in the project are maintained in the "metadata template file" (mtf), in the Metadata Specification link shown below.</p> <p>2. For each metadata parameter as appropriate, a list of approved values and their definitions is compiled by manual methods, based on input from experts at the ship operating institutions SIO, WHOI and LDEO, as well as from a controlled vocabulary in use with the Site Survey Data Bank for the Integrated Ocean Drilling Program.</p>
<b>Output products</b>	<p>Controlled Vocabulary Dictionary for shipboard and submergence data with approximately 700 entries, published in the Dictionaries link below. To illustrate the approach, sections of the dictionary are presented:</p> <p>MMI_CanonicalADO_Subject,"current-profiler_adcp","Acoustic doppler current profiler"  MMI_CanonicalADO_Subject,"depth_current-version","Best available echo sounding or center beam depth time series"  MMI_CanonicalADO_Subject,"depth_raw-archive","Original raw version of depth data"  MMI_CanonicalADO_Subject,"meteorology","High resolution meteorological data from various shipboard sensors"  MMI_CanonicalADO_Subject,"multibeam_bathymetry_current-version","Best available multibeam swath data"  MMI_CanonicalADO_Subject,"multibeam_bathymetry_current-version_sdfile","Best available multibeam visualization data"  MMI_CanonicalADO_Subject,"navigation_current-version","Best available navigation data"  MMI_CanonicalADO_Subject,"sidescan","Towed sidescan acoustic backscatter"  MMI_CanonicalADO_Subject,"SeismicMultiChannel","Multi channel seismic reflection deep penetration data with streamer and low frequency source such as air gun"</p> <p>SIOGDC_Cruise_Vessel,"R/V ALEXANDER AGASSIZ","SIO 1962-1977"  SIOGDC_Cruise_Vessel,"R/V ARGO","SIO 1960-1970"  SIOGDC_Cruise_Vessel,"R/V ATLANTIS","WHOI AGOR-25 1996-"  SIOGDC_Cruise_Vessel,"R/V ATLANTIS II","WHOI 1963-1997"</p> <p>SIOGDC_Instrumentation_Model,"SB2000 multibeam","second generation SeaBeam 121 beam system on Melville 1992-2007"  SIOGDC_Instrumentation_Model,"SB2112 multibeam","third generation SeaBeam 141 beam system on Revelle 1996-2000, also on Knorr 1994-"  SIOGDC_Instrumentation_Model,"EM120 multibeam","Kongsberg Simrad 191 beam system on Revelle 2001-"</p> <p>SIOGDC_Instrumentation_Method,"hull mounted sonar","permanent installation on hull"  SIOGDC_Instrumentation_Method,"pole mounted sonar","temporary installation on pole"  SIOGDC_Instrumentation_Method,"towed streamer","sensors towed behind vessel in long streamer, usually acoustic"  SIOGDC_Instrumentation_Method,"towed magnetometer","magnetometer towed behind vessel on long cable"</p> <p>SIOGDC_Samples_Device,"biology sampler - colonization substrate"  SIOGDC_Samples_Device,"biology sampler - crab trap","sampler for crabs"  SIOGDC_Samples_Device,"biology sampler - incubation chamber","sampler for long term deployment"  SIOGDC_Samples_Device,"dredge - rock","rock dredge"  SIOGDC_Samples_Device,"fluid sampler - gas tight","sampler that securely seals to preserve fluid for laboratory gas analysis"</p> <p>SIOGDC_Documentation_Type,"Bridge Log","ships officers version of shipboard events"  SIOGDC_Documentation_Type,"Citation","citation of publication"  SIOGDC_Documentation_Type,"Computer listings","record of computer run or data listing"</p>

	<p>SIOGDC_Documentation_Type,"Cruise Report","overall cruise report, with ports, dates, participants, track, profiles as available"</p> <p>SIOGDC_Submergence_Vehicle,"ABE","WHOI Autonomous Benthic Explorer with bathymetry, magnetics and other capabilities, launched 1995, replaced by Sentry 2007"</p> <p>SIOGDC_Submergence_Vehicle,"ALVIN","WHOI Alvin manned submersible, 1964"</p> <p>SIOGDC_Submergence_Vehicle,"DSL120a","WHOI deep submergence towed survey system, with sidescan and bathymetry, 4.5 KHz subbottom profiler, fiber optic cable, since 2001, will cease to be a formal part of the NDSF in 2007"</p> <p>SIOGDC_Submergence_Vehicle,"JASON2","WHOI deep submergence survey system ROV with thrusters and multiple imaging and acoustic sensors, replacing original JASON in 2002"</p> <p>SIOGDC_Submergence_Vehicle,"DeepTow","SIO deep tow vehicle, multiple cameras, sonars, other sensors, 1964-2002"</p> <p>SIOGDC_Submergence_NavigationType,"LongBaseLine","vehicle navigation from acoustic transponders on sea floor"</p> <p>SIOGDC_Submergence_NavigationType,"DP","vehicle navigation with dynamic positioning"</p> <p>SIOGDC_Submergence_NavigationType,"DVLNav","vehicle navigation from Doppler velocity system"</p> <p>SIOGDC_Submergence_NavigationType,"USBL","vehicle navigation from ultra short baseline system on hull of vessel"</p>
<p><b>Steps in approach</b></p>	<p><b>Preparation</b></p> <ul style="list-style-type: none"> <li>• Begin with a list of formal metadata parameter names and their definitions from the Metadata Specification (link below). The parameter definition is often helpful in seeding the value definitions needed in the dictionary.</li> <li>• For each parameter as appropriate, create a record in the dictionary for each approved value, and insert a brief definition for the value with a text editor. The collection manager or archivist for each ship operating institution is generally responsible for gathering candidate values and selecting authorized values. One empirical approach is to use a database to look at all the posted values for a specific parameter across a collection. Naming conventions change over decades and across institutions. This phase takes some time, and involves iterative collaboration with other ship operating institutions and collection managers.</li> <li>• Post the dictionary online so it can be used by various applications and by other institutions.</li> <li>• New records are easily added to the cvdict as the project evolves.</li> </ul> <p><b>Operation</b></p> <ul style="list-style-type: none"> <li>• During the staging phase of digital library collection building, the SIOExplorer auto-harvesting scheme uses the cvdict as it makes determinations of data types based on directories and filenames. The scheme also auto-populates metadata for each object, drawing values from the cvdict as appropriate. When candidate metadata values are extracted from data files or other sources, they are compared to approved values and flagged for manual intervention when conflicts arise. In practice, perhaps 5 percent of metadata extracted from historical sources proves to be in error, either from mis-spellings, unapproved synonyms, or misconceptions typed on a keypunch on a rolling ship.</li> <li>• Researchers and students conduct search queries with an interface that uses drop-downs from the cvdict. Behind the scenes, the interface creates SQL that is guaranteed to find a pattern match in the collection.</li> </ul> <p><b>Further usage of results</b></p> <ul style="list-style-type: none"> <li>• Real time data metadata generation approaches can make use of the cvdict to insert correct values at acquisition time during the cruise, rather than in the laboratory during final archiving.</li> </ul>
<p><b>Results to date</b></p>	<p>Working system in operation at SIO and WHOI.</p>
<p><b>Future goals</b></p>	<p>The dictionary is available for use with LDEO RIDGE and MARGINS database activities, and other ship operating institutions.</p>
<p><b>Lessons learned</b></p>	<p>When a practical definition is included for each value, the barriers from acronyms and domain specific jargon can be broken down.</p>
<p><b>Re-usability</b></p>	<p>While this project happens to be based on the mtf metadata specification, the controlled vocabulary dictionary is easily applied to other projects. The only requirement is that the first field in each record is the formal parameter name used in a project. The only technical skill required for updating the dictionary is the use of a text editor.</p>
<p><b>Presentations</b></p>	<p>"SIOExplorer" presentation for International Data Exchange Workshop, Kiel, May 2007  <a href="http://gdccoll.ucsd.edu:8080/digarch/about-project/presentations/kielsioexplorer">http://gdccoll.ucsd.edu:8080/digarch/about-project/presentations/kielsioexplorer</a>  "The Role of Controlled Vocabularies in Digital Archiving" AGU abstract, December 2006  <a href="http://gdccoll.ucsd.edu:8080/digarch/about-project/presentations/agu2006abstract">http://gdccoll.ucsd.edu:8080/digarch/about-project/presentations/agu2006abstract</a>  Digital archiving and preservation project background information  <a href="http://gdc.ucsd.edu:8080/digarch/about-project/presentations/">http://gdc.ucsd.edu:8080/digarch/about-project/presentations/</a></p>

<b>Reports</b>	
<b>Tools</b>	Next-generation SIOExplorer search interface under construction
<b>Metadata specification</b>	<a href="http://siox.sdsc.edu/metadata/mtf/current/MTF_current">http://siox.sdsc.edu/metadata/mtf/current/MTF_current</a>
<b>Standards</b>	SIO/SDSC mtf approach with canonical collection, ado and native blocks, although cvdict is independent
<b>Format</b>	Plain text
<b>Data products</b>	Revised metadata files for the entire SIOExplorer collection for the next version of the digital library, <a href="http://siox.sdsc.edu">http://siox.sdsc.edu</a>
<b>Dictionaries</b>	
<b>Controlled Vocabularies</b>	The cvdict combines the role of controlled vocabulary (for values) and dictionary (definition for each value) <a href="http://siox.sdsc.edu/metadata/CV/current/CV_current">http://siox.sdsc.edu/metadata/CV/current/CV_current</a>
<b>Ontologies</b>	
<b>Protocols</b>	
<b>Schemas</b>	
<b>Web services</b>	
<b>Other products</b>	
<b>Funding sources</b>	"DIGARCH" NSF/Library of Congress CISE IIS 0455998 Multi-institution Testbed for Scalable Digital Archiving
<b>Start date</b>	2005-06-01
<b>End date</b>	2008-05-30
<b>Sustainability plan</b>	Geological Data Center, SIO responsibility for ongoing effort
<b>Case Study contributor</b>	Stephen Miller
<b>Case Study date</b>	2007-06-01