

Coastal and Marine Spatial Planning Discovery Vocabularies Workshop

A Joint WHOI/USGS/NOAA Workshop

12/1/2010-12/3/2010

**Workshop Report
February 18, 2011**

**Woods Hole Oceanographic Institution
Woods Hole, Massachusetts**



A note on the formulation of this report:

This report was a collaborative effort produced largely during the meeting with the use of an on-line editing tool called iEtherPad. All participants had access to the document and made contributions during and after the workshop. Breakout groups and individuals took responsibility for sections of the report, which encapsulate what took place during the workshop and the conclusions formed. The WHOI facilitators and workshop steering committee edited and organized the content (as an MSWORD document), but it was very much a collaborative writing effort.

More information on the workshop can be found at:

<http://www.who.edu/sites/cmsspworkshop>

Questions and comments regarding this report can be directed to:

Andrew Maffei, Meeting Leader: amaffei@who.edu

Patricia White, Meeting Facilitator: pwhite@who.edu

Table of Contents

EXECUTIVE SUMMARY	2
Recommendations	3
INTRODUCTION	4
Workshop Purpose	4
Workshop Background.....	4
Overarching Principles	5
BREAKOUT GROUPS: REPORTS AND RECOMMENDATIONS.....	5
Breakout Group #1.....	6
TOPIC #1. Roles and Responsibilities Related to Discovery Vocabularies.....	6
TOPIC #2. A Suggested Process for Adopting Mature and Valid Vocabularies.....	8
TOPIC #3. Testing the Usefulness/Applicability of Criteria in #2 above	9
TOPIC #4. Criteria for Evaluating Use Aspects of Vocabularies.....	9
Breakout Group #2.....	10
TOPIC #5. Suggestions to the National Ocean Council	10
CONCLUSIONS.....	13
Post-workshop Activities	13
Strategy going forward.....	13
APPENDICES	15
Appendix A: Workshop Participants.....	15
Appendix B: Survey Results: Crosswalks.....	16
Appendix C: Survey Results: Repositories	16
C.1. Data Repositories.....	18
C.2. Vocabulary Repositories.....	20
Appendix D: Survey Results: Metadata Standards	21
Appendix E: Survey Results: Tools	24
Appendix F: Survey Results: Vocabularies	27
Appendix G: Descriptions of Vocabulary-Related Resources in other Projects.....	31
Appendix H: Criteria for Evaluation Intrinsic Aspects of Vocabularies.....	33
Appendix I: Acronymns.....	35
Appendix J: Agenda.....	36

EXECUTIVE SUMMARY

Discovery vocabularies for coastal and marine spatial planning were the focus of a three-day USGS/NOAA interagency workshop that was hosted by Woods Hole Oceanographic Institution (WHOI) on Dec. 1-3, 2010. Peter Fox of Rensselaer Polytechnic Institute (RPI) provided technical leadership for the workshop. Andrew Maffei of WHOI was the workshop leader. Patricia White served as facilitator. Workshop participants were mostly from USGS, NOAA, and WHOI, with additional invited participants from the Gulf of Maine Research Institute, Lamont-Doherty Earth Observatory, the Marine Biological Laboratory, and the Oregon Coastal Management Program.

In response to the National Ocean Policy, all Federal agencies are required to participate in the process of coastal and marine spatial planning (CMSP). A critical challenge will be populating data catalogs with appropriate metadata so that information assets can be more readily discovered, evaluated for suitability, integrated with data provided by other agencies, and used more effectively by planning bodies, resource managers and stakeholders across the country. The workshop was held to assist USGS and NOAA in sharing best practices and metadata tools in order to reduce the burden on both government and non-government scientific programs and offices providing information for CMSP.

The workshop began with a keynote address on “National Ocean Policy, CMSP Framework and its Data Management Requirements” presented by Betsy Nicholson and Tony LaVoi of the NOAA Coastal Services Center. In addition to clarifying the requirements of CMSP, LaVoi expressed interest in taking ideas from the workshop back to the interagency group that is designing a national information management system and data portal for CMSP.

Peter Fox presented a second keynote address entitled, “Interoperable Data Catalogs and Discovery Vocabularies”. This talk introduced several ideas that were major themes during the rest of the workshop. Workshop participants realized that they were dealing with *curation*, which requires a focus on how the CMSP community will look for data, rather than on the characteristics of the data itself. Much of our data is accessible via web services or databases using traditional web-based query interfaces. Making these data adequately available will require semantically aware or faceted search capabilities, employing vocabularies that are more than simple lists of terms. Finally, governance will be needed to collaboratively evaluate, define, and maintain the vocabularies over time to meet emerging requirements.

Most of the workshop was spent in discussing and refining (1) criteria and a process for evaluating vocabularies and (2) vocabulary and metadata suggestions for Tony LaVoi to take to the interagency group. Workshop participants agreed to test the process for evaluating vocabularies and drafted the following recommendations. Lists of resources, collected from participants and pertinent to CMSP vocabulary use and development appear in appendices at the end of this report.

More information on the workshop can be found at:
<http://www.whoi.edu/sites/cmsspworkshop>.

Recommendations

The workshop participants make the following recommendations.

- Recommendation 1:** We recommend that the CMSP information system define a clear set of roles and responsibilities for people who work with CMSP discovery vocabularies in the realm of metadata handling, vocabulary development, and other related tasks.
- Recommendation 2:** We recommend that the CMSP information system include a standard process for adopting as ‘community-accepted’ mature and valid discovery vocabularies for use in CMSP efforts and that the process include testing with national and regional data repositories and discovery systems.
- Recommendation 3:** We recommend that a subset of our workshop attendees be empowered to establish community presence within the Marine Metadata Initiative (MMI) to provide for ongoing community interaction and knowledge management related to the evaluation and recommendation of CMSP discovery vocabularies.
- Recommendation 4:** We recommend that the CMSP information system include a strategy for educating and supporting those responsible for developing and managing repositories holding CMSP-related data to adopt mature and valid discovery vocabularies.
- Recommendation 5:** We recommend that the CMSP information system include best practices, documentation, cookbooks, and/or other explanatory materials to support understanding of the principles and the application of discovery vocabularies (generally and in the specific context of CMSP).

INTRODUCTION

Workshop Purpose

A three-day NOAA/USGS interagency workshop hosted by Woods Hole Oceanographic Institution (WHOI) with technical leadership from Rensselaer Polytechnic Institute (RPI) was held on December 1-3, 2010 to address these needs of the future national process of coastal and marine spatial planning (CMSP):

- Choosing shared standards for metadata content for data catalogs that will support the national process for coastal and marine spatial planning, including shared vocabularies;
- Cataloging vocabularies, ontologies and crosswalks that can be used to improve interoperability;
- Sharing information about tested tools for creating metadata for new and existing catalog services, and providing guidance for using these tools and services;
- Establishing communication mechanisms for future consultation among workshop participants and the larger CMSP community.

Workshop Background

In response to the National Ocean Policy, all Federal agencies are required to participate in the process of coastal and marine spatial planning (CMSP). The Federal Government is developing a national information management system that will provide access to appropriate data from many sources, including agencies from all government levels and the academic research community. In addition, regional planning bodies may have their own information systems and regional data sources. For this distributed system to work well, a critical challenge will be populating data catalogs with appropriate metadata so that information assets can be readily discovered, evaluated for suitability, integrated with data provided by other agencies, and then used by planning bodies, resource managers and stakeholders across the country. By sharing best practices and metadata tools, WHOI, USGS, NOAA and others can reduce the burden on the scientific programs and offices providing information for CMSP.

Overarching Principles

Data and metadata must be properly cared for in order for the data to continue its usefulness in the long term. Proper **curation** of data and metadata implies that careful attention will be given to the preservation and maintenance of both a data set and its accompanying metadata. The CMSP community will need to educate and encourage data providers on curation of data sets.

Data of interest to the CMSP community will come from a wide variety of scientific and non-scientific disciplines and sources. Different disciplines currently use and will continue to create vocabularies that are specific to users and needs within those disciplines. As such, the CMSP community must identify methodologies for choosing, adopting and integrating a number of vocabularies which, when taken together, are sufficient for use in metadata that is useful to CMSP. The need for this variety of vocabularies implies a further need to investigate and support the development of crosswalks and ontologies so that data sets from different sources can be more effectively integrated and understood. As vocabularies are added, updated, and expanded, crosswalks and their mappings to these vocabularies will need to be maintained in order to remain useful.

For all of this to be sustainable, there must be an appropriate level of **governance**. Governance includes technical activities such as maintenance issues (e.g., versioning, audit trails and standards), as well as moderating vocabulary content state and forums for discussion. The governance structure must be stable enough to be authoritative, inclusive of related stakeholders, and flexible enough to develop with the needs of the community.

Encouraging and supporting the use of controlled vocabularies for designated metadata will ensure that data can be discovered, evaluated, integrated and used in systematic and repeatable ways. Both data providers and data users will benefit by the adoption of controlled vocabularies.

BREAKOUT GROUPS: REPORTS AND RECOMMENDATIONS

Breakout Group #1

The charge to breakout group #1 was to consider how best to evaluate vocabularies for validity and usefulness. This group addressed these four primary topics in a relatively short period of time:

1. Roles and responsibilities related to discovery vocabularies,
2. A suggested process for adopting mature and valid discovery vocabularies,
3. Testing the usefulness/applicability of criteria in #2 with existing CMSP repositories, and
4. Criteria for evaluating use aspects of vocabularies.

These resulting topics and related conclusions are described below.

TOPIC #1. Roles and Responsibilities Related to Discovery Vocabularies

In order to accomplish the goals of the CMSP, individuals will need to adopt specific roles (and their associated responsibilities). It is unclear whether anyone participating in CMSP will be directly paid by a CMSP organization. What will help make CMSP successful is effective cooperation and coordination among groups like FGDC, MMI, IOOS, for example, that have active communities of practice that include data providers, collection managers, portal managers, and other stakeholders from such organizations.

The definitions of the following roles are fluid and deliberately generalized. In the real world, chains of custody and levels of responsibility for data sets assume different players often with multiple responsibilities and different titles. Data providers in some communities turn their data over to a collection manager almost immediately upon return from a field activity. In some communities, these data are well documented from the instrumentation to project descriptions and processing notes. In other communities, it may be years before data has been fully analyzed and prepared for release. Data may be acquired from additional parties and then processed, combined, compared, etc. before a data set of value to the CMSP effort is available for use. Many data providers are contributing their raw data to a number of collections for integration and use.

The time and method of assigning – or expanding – appropriate metadata (and application of associated vocabularies) can occur at every point in the chain of custody. Having the data provider produce the most complete metadata record possible is the preferred approach. There is also not always a linear path of custody and metadata creation.

Draft Roles and Responsibilities related to discovery vocabulary management:

- **Data Producers** select appropriate vocabularies, (e.g., locational, thematic, administration/organizational, etc.) for essential elements of their data and metadata that have community-accepted standards. They have the responsibility to recognize the distinction between usage vocabularies that accurately describe particular data fields and the more general discovery vocabularies that are designed to help others find the data.
- **Collection Managers** identify essential elements that can enhance access when controlled vocabularies are available and used. They can then work with *data producers* to discover and adopt stable/authoritative vocabularies for those elements.
 - If CMSP specific terms/elements/vocabularies are available, the collection manager might be able to apply them.
 - Assistance could be made available to both *data producers* and *collection managers* in applying consistent vocabularies after the initial creation of metadata.
- **Regional/Thematic Data Portal Managers** expose the metadata that is available for data so that communities can see how usable it is to accomplish discovery, evaluation, integration and use of data.

- If, down the chain, *data producers* are using accepted vocabularies, this can be seen in this exposure.
- Some regions/themes might not have data/metadata portals, so exposure of metadata may fall to *data producers* or their affiliated *collection managers*.
- **Data Providing Organizations** will need tools that can help *data producers* and other creators of metadata select the right, (i.e., accepted by the community as ‘authoritative’ for the purpose at hand), vocabularies and metadata standards for describing their data. Some people want to be told explicitly what they should use. Others are going to want to evaluate their options for their specific needs.
- **National Information Managers** are aware of the basic content and intended application of ‘community-accepted’ vocabularies, ontologies and crosswalks; support good data management practices and identify high priority areas for standardization to improve the discovery, evaluation, integration and use of data archiving.
- **CMSP Advisory Groups** need to include expertise in their membership that can recognize the need to use mature and valid vocabularies, ontologies, and crosswalks. “I don’t see the kind of information I need” may indicate a vocabulary that needs improvement rather than missing information.
- **Software Tool Developers** should be engaged, educated, and advised about CMSP vocabulary usage so their work can more effectively serve CMSP *contributors*.
- **Vocabulary Developers and Maintainers** should be engaged, educated, and advised. Their work can become more useful to CMSP by:
 - Making their vocabularies available in an Internet-accessible form.
 - Evaluating vocabularies using the criteria expressed later in this report.
 - Capturing the results of these evaluations in an accessible knowledge base.

Note: Climate and Forecast (CF) is an example of this process with effective communication (see listing in Appendix D).

- **"Data Wranglers"** can assist in data discovery by contributing domain knowledge to the development of metadata vocabularies and ontologies. These specialists are called 'Data Scientists' in some other communities.

The CMSP initiative needs to find ways to engage these stakeholders constructively. Although technology such as software tools, online manuals and websites can be of assistance, face-to-face interactions and training are also necessary.

Recommendation 1: We recommend that the CMSP information system define a clear set of roles and responsibilities for people who work with CMSP discovery vocabularies in the realm of metadata handling, vocabulary development, and other related tasks.

TOPIC #2. A Suggested Process for Adopting Mature and Valid Vocabularies

Goal:

Develop the outline and initial elements of a process for identifying, evaluating and adopting mature and valid vocabularies for use in the area of CMSP and related scientific and non-scientific domains.

Draft Process:

1. Identify the type of data.
2. Identify the essential elements for those data to support discovery, evaluation, integration, and use in the CMSP context.
3. Identify the need for specific vocabularies that apply to Essential Elements.
 - a. What is the requirement driving that need? For example, for time sequence data, the need is comparability to other data sets temporal attributes.
 - b. Which elements are amenable to the use of a vocabulary?
 - c. For which elements is “no vocabulary” preferable in order to capture information that is unique?
 - d. Which elements should use a community-accepted standard (e.g. date which has an ISO standard defined)?
4. If it is determined that a controlled vocabulary is desirable, identify appropriate existing vocabularies and apply the Criteria for Evaluating Use Aspects of Vocabularies (see Topic #4 below). How well does each vocabulary accomplish its intended purpose?
5. Document the vocabularies that have been adopted for your data so that others can use your metadata.
6. Periodically check for vocabulary updates and implement new terms.
7. Periodically audit your use of the selected vocabulary. Identify new needs (potentially from new disciplines) that are not being addressed by the vocabularies you have selected and are using.
8. If you so choose, periodically extend the metadata in your collection with current terminology (i.e. Retrospective Cataloging). If possible, alert the data provider when extending the metadata they provided with newly available terms.
9. Communicate with the vocabulary provider to contribute additional terms and term relationships or suggest modifications.

Recommendation 2: We recommend that the CMSP information system include a standard process for adopting as ‘community-accepted’ mature and valid discovery vocabularies for use in CMSP efforts and that the process include testing with national and regional data repositories and discovery systems.

TOPIC #3. Testing the Usefulness/Applicability of Criteria

After some discussion it was decided that the process and criteria developed at this workshop for adopting mature and valid vocabularies and ontologies needs to initially be tested by applying them to a subset of data that are of high priority for the CMSP process. The results of these tests, as well as ensuing recommendations for the adoption of specific vocabularies or for changes in the process and criteria, should be maintained in an accessible location. Ideally, a community of practice for CMSP discovery metadata would continue this work, maintaining a set of recommended vocabularies and refining the process and criteria for evaluation and selection.

TOPIC #4. Criteria for Evaluating Use Aspects of Vocabularies

(This section is drawn from recent experience in jointly designing cruise-level metadata for the Extended Continental Shelf (ECS) and Rolling Deck to Repository (R2R) programs, and draws from the Marine Metadata Interoperability (MMI) Guides. <<http://marinemetadata.org/guides/vocabs/cvchooseimplement>>.)

(* Indicates criteria that are potential disqualifications of candidate vocabularies or minimum requirements.)

Draft Criteria:

1. * Is the vocabulary available online, in a programmatically accessible XML-based format, with HTTP-style URIs?
2. * Is the vocabulary hosted in a stable namespace with a known technical contact?
3. Is the vocabulary explicitly versioned, and are earlier versions available online?
4. Is the vocabulary designed/suitable primarily for Discovery (Search) or for Use (Markup)? <<http://marinemetadadata.org/guides/vocabs/voctypes/cvusagevsdisc>>
5. Is the vocabulary connected to a larger context? Does it link explicitly to upper-level, lower-level and/or peer vocabularies?
6. Does the vocabulary make a clean distinction between types and instances?
7. Does the vocabulary have a governance structure? Is there a well-documented system for suggesting new terms and having them adopted for the larger community? (Email request to the author/maintainer? Formal committee/review structure?)
8. Is the vocabulary “flat” or hierarchical? If hierarchical, is it extensible to new levels?
9. Do mappings exist to other vocabularies? Are they maintained?
10. Do all vocabulary terms have definitions? Are they maintained? Are they promoted/required by the governance structure and terms of use?
11. Does the vocabulary have overview documentation and a cookbook/user guide? Is the documentation contained in the governance structure and maintained?
12. Have one or more science/disciplinary communities approved and/or adopted the vocabulary?

Breakout Group #2

The charge to breakout group #2 was to document advice and ideas mentioned by workshop participants, based on their career experience, and to pass this along to the National Ocean Council group that is designing the CMSP national information management system (NOC NIMS).

TOPIC #5. Suggestions for development of the CMSP National Information Management System (NOC NIMS)

The following 7 suggestions are offered for consideration by those charged with developing the NOC NIMS. In addition, we urge system designers to learn from the work of the Long Term Ecological Research (LTER) Network and other communities that have previously discussed many of the concepts described here (see Appendix G). LTER and other research groups with past experience in this area should be invited to future discussions of these suggestions.

SUGGESTION #1: Define a relatively small number of key top-level concepts needed to enable smart and faceted discovery searches for CMSP information (see Peter Fox presentation on "Interoperable Data Catalogs and Discovery Vocabularies" at <http://www.who.edu/page.do?pid=50215>).

- A faceted search requires a first layer-controlled vocabulary.
- A combination of structured and unstructured searches can use a vocabulary in RDF form to suggest concepts based on words typed into a text box. For an example, see how Amazon.com's shopping webpages operate.
- The need to search by geography, time and theme will be common across all regions.
- Maps are not always the best starting point in searches; street address, zip code or place name are often better. Place name searches depend on vocabularies:
 - marine gazetteers
 - "geonames" <<http://www.geonames.org/>> which has an expanded version with both places and terms including those from the National Geospatial-Intelligence Agency, the U.S. Board on Geographic Names, and the USGS Geographic Names Information System
 - a place name thesaurus to provide synonyms
- For transparency, it is important to ensure that the public can easily find CMSP data using simple search engines (Google, Bing). Use of controlled vocabularies in metadata tags will make such unstructured searches more effective.

SUGGESTION #2: Design the information system to enable access to *authoritative* sources of data, at whatever geographic or institutional scale.

- If the authoritative source is at a regional institution, design the system to draw from that source and not some adaptation of that source.
- To achieve this goal, pathways into the system need to allow for all levels of possible data contributors, including sources that are sub-national or sub-regional.

- Include tracking the identity of contributors of data or information. System users need to know who is providing the data.

SUGGESTION #3: Be cautious to avoid making the requirements for data sharing so onerous that small data providers opt out due to constrained resources. (i.e. Make it ‘easy to do the right thing’.)

- In setting standards for vocabulary or technology, consider the need to provide assistance to data providers who do not have the capacity to meet those standards on their own.
- Consider providing a hosted service that some of the smaller data providers can contribute data and metadata to.
- Consider sending data experts to visit small labs to help providers build capacity to provide data and rich associated metadata using controlled vocabularies.

SUGGESTION #4: A common infrastructure is needed to standardize vocabularies and to manage change -- some body facilitating the development and evolution of vocabularies so CMSP information will have metadata sufficient to be discovered, evaluated, integrated, and used.

- Leverage the vocabulary hosting and tools of the of the Marine Metadata Interoperability project.
- Vocabularies need to be extensible so they can evolve.
- Authentication and ID management is necessary to contribute and modify/annotate data and metadata.
 - Don't want a free-for-all on editing vocabularies
 - Do want to be able to have notations that can improve discovery, evaluation, integration, and use.
 - If there were a term or concept that someone would like to suggest, the community that manages the vocabulary would want to hear from them.
 - Accountability is required: One needs to attribute changes, when they occurred, and why (as with management/versioning used in software systems).
 - Quality control, auditing and authorization are issues.

SUGGESTION #5: Two-way communication will be needed at various levels - between information producers, collection managers, portal managers, end users, and the like. Feedback needs to get back to someone that can actually make a change or address the issue raised.

SUGGESTION #6: The degree of federation of the information system affects the choice of vocabularies. To the extent possible, vocabularies for discovery should occur close to the source of the data.

- The appropriate level of detail depends on what community you are serving: different individual use cases could imply different interfaces and vocabularies.
 - Although there is some homogeneity across a CMSP region, even at that level it is not complete.
 - CMSP managers might rely more on socio-economic and regulatory concepts than scientific concepts.
 - Different sources of data (commercial, oceanographic, geological) have different vocabularies

- If a vocabulary existed at the national level, this would be helpful so that external catalogs could offer crosswalks, mapping their existing vocabularies to the national vocabulary.
 - National level vocabulary terms should never replace local vocabulary terms. Local terms should, instead, map to more global terms.
 - We need to start at the "concept" level before getting to the physical schemas.
 - This suggestion is an attempt to provide vocabulary crosswalks between different communities (data providers and consumers) to help with data discovery, recognizing that vocabulary changes across communities and over time.
 - Consider the use of rules-based faceted search that is capable of shifting to a different crosswalk vocabulary based on prior terms used.
 - Note that software tools already exist to facilitate these crosswalks.

SUGGESTION #7: The NOC could approach the Geospatial Platform with the overall concept of what it is trying to do to see if these could be defined as common services that could be funded broadly at the national level, cutting the required investment by the NOC by leveraging the project for all domains of the government.

- This is a platform issue with vertical markets (pillars) that are logically defined where partners will get together to work out how to solve problems together.
- The CMSP could be seen as a segment. Is there a "segment" for discovery that could be applied broadly to other issues?
- Draws on the Federal Enterprise Architecture concepts.

CONCLUSIONS

Post-workshop Activities

At the end of the workshop we asked participants if they would be willing to test the application of the discovery vocabulary process described above, using the vocabulary criteria also described above, in some aspect of their ongoing work. The following volunteers offered to take on this and related tasks.

1. USGS Coastal and Marine Geology Knowledge Management Team will test the process with a selected vocabulary.
2. The Sea Ice work that Peter Fox is leading has completed vocabulary evaluation and selection and the results will be contributed to this CMSP effort.
3. Bob Arko has applied the process to about a dozen vocabularies and provided feedback to the vocabulary authors.
4. Anne Ball, NOAA Coastal Services Center, will talk to the socio-economic data people at NOAA to see what elements they look for and whether there are vocabularies available to meet their needs. (See Suggestion #6 above.) Also, her group is testing the CIESIN list of Socio-Economic Keyword (see entry in Appendix F) to see how well they work as part of this effort.
5. Viv Hutchison (USGS) will test the process with some biological vocabularies for the NBII program (see Appendix F.).

It will be important for groups to report their experiences with the vocabulary testing process even in cases where the process 'did not work'. In order to carry out consistent testing it will be important to have a list of top priority data sets for CMSP which also expose the broadest range of 'classes' of vocabulary 'issues'. Top priority data sets may include those with data on benthic habitat, bathymetry, and oblique shorezone characterization (i.e. cliffs, beach, etc.).

Recommendation 3: We recommend that a subset of our workshop attendees be empowered to establish community presence within MMI to provide for ongoing community interaction and knowledge management related to the evaluation and recommendation of CMSP discovery vocabularies.

Strategy Going Forward

The implementation of vocabularies in the CMSP framework will need to follow existing standards and conventions. The metadata needed to support CMSP will come from many sources and from data providers with differing abilities to meet additional requirements. Also, the metadata needed to support CMSP will be used in other programs and projects so it is important not to add unusual or unnecessary requirements to data providers' existing workloads. Doing so would restrain data providers from meeting the needs of CMSP.

In order to move CMSP forward, the national information system needs to investigate and adopt metadata standards, vocabularies, crosswalks and ontologies that are appropriate for use in the metadata for those data of interest. Educational materials will be needed to help data providers understand how to implement these standards and vocabularies. Portals or data systems developed for CMSP will need to integrate the crosswalks and ontologies as appropriate.

Importance of Education

Education on the incorporation of vocabularies into metadata will be necessary. Educational materials will need to be available to a wide audience and provide sufficient information for data providers to easily include terms of importance to CMSP. On-line training and resources will probably be the most beneficial as these materials can be accessed at the data providers' convenience.

Recommendation 4: We recommend that the CMSP information system include a strategy for educating and supporting repositories with CMSP-related data in adopting mature and valid discovery vocabularies.

Engaging Communities outside of the Earth Sciences (See Suggestion #6 under Topic 5 above.)

CMSP should investigate vocabularies and metadata standards being used in communities of interest that are outside the area of the earth sciences. It will be useful to engage with organizations such as the Center for International Earth Science Information Network (CIESIN) and others who have experience with socio-economics and related disciplines.

Recommendation 5: We recommend that the CMSP information system include best practices, documentation, cookbooks, and/or other explanatory materials to support understanding of the principles and the application of discovery vocabularies (generally and in the specific context of CMSP).

Sustaining the CMSP Community

Providing a community presence via the websites of organizations such as the Marine Metadata Initiative (MMI) and the Federal Geographic Data Committee (FGDC) would be useful for communicating information about vocabulary use for CMSP.

Appendix A

WORKSHOP PARTICIPANTS

(* Indicates member of steering committee.)

	Participant	Institution
1	Alan Allwardt	USGS
2	Bob Arko	Lamont Doherty
3	Anne Ball*	NOAA/NOS/CSC
4	Eric Bridger	GMRI
5	Cynthia Chandler	WHOI/MCG
6	VeeAnn Cross	USGS/CMG/WHSC
7	Peter Fox*	RPI/TWC
8	Art Gaylord*	WHOI/CIS
9	Philip Goldstein	USGS/NBII/OBIS-USA
10	David Govoni	USGS/EI
11	Tanya Haddad	State Employee, Oregon
12	Richard Huffine	USGS/GIO/Library
13	Vivian B Hutchison*	USGS
14	Fran Lightsom*	USGS/CMG/WHSC
15	Andy Maffei*	WHOI/CIS/OIWG
16	James Manning	NEFSC
17	Daniel Martin	NOAA/NOS/CSC
18	Kathy Martinolich*	NOAA/NCDDC
19	Betsy Nicholson	NOAA
20	Peter Schweitzer	USGS/Geology/Eastern
21	Sharon Shin	FGDC
22	Derrick, Snowden	NOAA/OAR
23	Trish White	WHOI

Appendix B

SURVEY RESULTS: CROSSWALKS

1. **Alexandria Digital Library crosswalks** <<http://www.alexandria.ucsb.edu/public-documents/metadata/crosswalks.html>> Crosswalks from the Alexandria Metadata Schema to Other Schemas
2. **EPA Network Exchange** <<http://www.epa.gov/networkg/index.html>> is the Exchange Network & Grant Program on the USEPA home page
3. **FGDC CSDGM to GCMD**
<http://gcmd.nasa.gov/Aboutus/standards/fgdc_to_dif.html> a mapping of fields from the FGDC Content Standard for Digital Geospatial Metadata to the GCMD DIF.
4. **FGDC CSDGM to ISO crosswalk, FGDC**
<http://www.fgdc.gov/metadata/documents/FGDC_Sections_v40.xls>
5. **GCMD mappings** <<http://gcmd.gsfc.nasa.gov/Aboutus/standards/>>
6. **Getty Research Institute compilation**
<http://www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html>
7. **ICAN ontology mappings** <<http://ican.science.oregonstate.edu/node/571>> several small ontologies have been developed within the ICAN tech team in order to demonstrate the potential utility to member projects
8. **MIT Libraries compilation of Metadata Mappings (Crosswalks)**
<<http://libraries.mit.edu/guides/subjects/metadata/mappings.html>>
9. **MMI compilation of Vocabulary Crosswalks**
<<http://marinemetadata.org/taxonomy/term/159>> Collection of links to individual crosswalks as well as other lists of crosswalks
10. **NERC Data Grid, BADC** <<http://ndg.nerc.ac.uk/>> The NERC Data Grid is a federated data infrastructure that delivers a variety of data-related services
11. **NOAA FGDC CSDGM/MARC21/Dublin Core**
<<http://coris.noaa.gov/data/examples/MetadataCrosswalk.pdf>>
12. **SeaDatanet** <<http://www.seadatanet.org/>> The SeaDataNet objective is to network and enhance currently existing infrastructures via a standardized system to manage, access and share data, information, products and knowledge originating from oceanographic fleets, new automatic observation systems and space sensors.
13. **SWEET to CF** <<http://iridl.ldeo.columbia.edu/ontologies/cf2sweet.owl>> classification of the MMI CF standard names by SWEET -- perl script translation of the cf.xls spreadsheet on the JPL website to RDF

14. **UK Office for Library and Information Networking compilation** - Mapping
between Metadata Formats
<<http://www.ukoln.ac.uk/metadata/interoperability/>>

Appendix C

SURVEY RESULTS: REPOSITORIES

(*Note: Numbers in parenthesis (x) indicate the number of people who listed that item as a resource in the survey.)

C1. Data Repositories

1. **Biological and Chemical Oceanography Data Management Office** (BCO-DMO), U.S. National Science Foundation <<http://osprey.bco-dmo.org/>> BCO-DMO works with PIs to manage data and results from research projects funded by the NSF Geosciences Directorate (GEO) Division of Ocean Sciences (OCE) Biological and Chemical Oceanography Sections and Office of Polar Programs (OPP) Antarctic Sciences (ANT). BCO-DMO manages and serves marine biogeochemical, ecological and oceanographic data and information developed in the course of scientific research and contributed by the originating investigators.
2. **CMG InfoBank**, USGS <<http://walrus.wr.usgs.gov/infobank/>> (3) A catalog of information about field activities of the USGS Coastal and Marine Geology Program, including some information about the resulting data and publications.
3. **Coastal and Marine Geology Program Internet Map Server** and GIS data, USGS <<http://coastalmap.marine.usgs.gov/>> contains links to current and past projects that have downloadable GIS data and metadata
4. **Data.Gov** <<http://www.data.gov/>>. (2) The purpose of *Data.gov* is to increase public access to high value, machine-readable datasets generated by the executive branch of the federal government.
5. **Ecowatch Catalog**, NOAA/NCDDC <<http://ecowatch.ncddc.noaa.gov/>> The purpose of Ecowatch is to provide a central point of access to coastal and oceanographic data and information. Ecowatch is a collection of services designed to support scientists and decision-makers by providing different ways to search for and visualize data.
6. **Geographic Names Information System**, USGS <<http://geonames.usgs.gov/>> Contains information about physical and cultural geographic features in the United States and associated areas, both current and historical (not including roads and highways). The database holds the Federally recognized name of each feature and defines the location of the feature by state, county, USGS topographic map, and geographic coordinates.
7. **Geoscience Data Catalog**, USGS <<http://geo-nsdi.er.usgs.gov/>> FGDC CSDGM metadata produced by the geologic division of USGS from 1995 through around 2002
8. **Geospatial OneStop** <<http://www.geodata.gov/>> (5) The US NSDI metadata catalog uses basic FGDC CSDGM metadata with ISO theme keywords.

Geospatial OneStop includes all types of geospatial data from all sectors within the USA. Will soon become part of Data.gov.

9. **Global Change Master Directory** <<http://gcmd.nasa.gov/>> (5) - NASA's Global Change Master Directory (GCMD) is a metadata catalog of Earth science data sets, services and tools.
10. **Marine Realms Information Bank**, USGS <<http://mrib.usgs.gov>> (2) A digital library providing faceted searches of free online scientific information, using a controlled vocabulary. MRIB also allows map-based geographic searches for information.
11. **Mineral Resources On-Line Spatial Data**, USGS <<http://tin.er.usgs.gov/metadata/>> listings of large national and global databases on mineral resource-related geology, geochemistry, geophysics, and mineral occurrence
12. **Mineral Resources On-Line Spatial Data Catalog**, USGS <<http://tin.er.usgs.gov/catalog/>> - Topical index of downloadable geospatial data including a metadata catalog, map interfaces, and specialized data of smaller regional and local studies on mineral resource-related geology, geochemistry, geophysics, and mineral occurrence
13. **National Geologic Map Database**, USGS <<http://ngmdb.usgs.gov/>> a bibliographic database containing thin, nonstandard metadata for geologic maps produced by USGS, state geological surveys, and academic institutions
14. **NBII Metadata Clearinghouse** <<http://metadata.nbii.gov/clearinghouse/>> (2) The NBII Clearinghouse is a powerful resource for scientists, allowing them to share and access information about important research in natural resources. With over 90,000 records from over 80 data providers, the possibilities for collaborations and data exchange are endless. Click here to read more about the NBII Clearinghouse
15. **Northeast Fisheries Science Center Oceanography Branch Data/Mapping Interface**, NOAA/NEFSC <<http://www.nefsc.noaa.gov/epd/ocean/MainPage/ioos.html>> provides access to water column, trawl survey, drifter, hull mounted sensor, moored time series, wind stress, and temperature and salinity anomaly data.
16. **NOS Data Explorer**, NOAA/NOS <<http://oceanservice.noaa.gov/dataexplorer/>> The NOS Data Explorer serves as a portal to obtain NOS spatial data, including bathymetry, coastal maps, environmental sensitivity index maps, aerial photographs, and more
17. **Oregon Coastal Atlas** <<http://www.coastalatlasc.net/search/>> This is the main access point for the archive of geospatial data available at the Oregon Coastal Atlas. Includes data accumulated as a result of grant activities conducted by a wide array of state agency and local jurisdiction conducted using NOAA CZM funds.

18. **Publications Warehouse**, USGS <<http://pubs.usgs.gov/>> (2) - Bibliographic records of USGS-authored journal articles, reports, and digital data releases. A Web service is available in the MODS format for integration with other catalogs.
19. **Science Topics directory**, USGS <<http://www.usgs.gov/science/>> (2) general topical browse index for all USGS scientific subjects, providing a selected subset of web-accessible resources to the public
20. **USGS Library Catalog** <<http://library.usgs.gov/>>

C2. Vocabulary Repositories

21. **BODC/NERC Vocabulary Server**
<http://www.bodc.ac.uk/products/web_services/vocab/>
22. **British Geological Survey vocabularies**
<<http://www.bgs.ac.uk/data/vocabularies/>>
23. **E-Science at RPI List of Ontologies** <<http://escience.rpi.edu/ontology/>>
24. **MMI Ontology Registry and Repository (2)**
<<http://marinemetadata.org/mmiorrusrman/>> A web-based application where ontology providers and users can collaborate to create, host, annotate, share, and use ontologies of relevance to the marine science community
25. **SWEET Semantic Web for Earth and Environmental Technology**
<<http://sweet.jpl.nasa.gov/>> - SWEET ontologies are written in the OWL ontology language and are publicly available. [SWEET 2.1](#) is highly modular with 6000 concepts in 200 separate ontologies.

Appendix D

SURVEY RESULTS: METADATA STANDARDS

1. **CF Conventions (2)** - NetCDF Metadata Climate Forecast and Conventions <<http://cf-pcmdi.llnl.gov/>>; CF Conventions Governance outlines how changes can be made. <<http://cf-pcmdi.llnl.gov/governance>>
2. **DarwinCore** <<http://rs.tdwg.org/dwc/>> A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. The Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, and samples, and related information.
3. **Dublin Core Metadata Element Set** <<http://www.dublincore.org/documents/dces/>> Dublin Core is a widely misunderstood metadata standard. The Dublin Core Metadata Element Set (DC- MES) is also known as Simple Dublin Core. Simple Dublin Core is a basic 15-element set designed to represent core features across all resource formats. It is standardized as ISO 15836-2003, ANSI/NISO Z39.85- 2007, and IETF RFC 5013. The Dublin Core Usage Guidelines sometimes suggest (but do not require) specific content guidelines or controlled vocabularies. Simple Dublin Core is widely known as the baseline metadata format required for all resources shared via OAI-PMH. Encoding of the DCMES in HTML <meta> tags was popular in the early days of search engines, but today most search engines prefer to weigh page text and linking patterns more heavily than page creator-supplied structured metadata. from: Glossary of Metadata standards by Jenn Riley, Devin Becker, Indiana University Libraries, Copyright 2009-2010 Jenn Riley <http://www.dlib.indiana.edu/~jenrile/metadatamap/seeingstandards_glossary_pamphlet.pdf>
4. **EML - Ecological Markup Language** <<http://knb.ecoinformatics.org/software/eml/eml-2.0.1/>>- EML grew out of early metadata efforts from the Ecological Society of America. It is an extremely detailed specification that is intended to support the description of any type of ecological information, including raw data, published research papers, rights information, and research protocols. EML is defined as a series of W3C XML Schemas, and can wrap data packages together with metadata. At the highest level, EML models four primary entities: datasets, literature, software, and protocols, although not all are always applicable or are required for use. from: Glossary of Metadata standards by Jenn Riley, Devin Becker, Indiana University Libraries, Copyright 2009-2010 Jenn Riley <http://www.dlib.indiana.edu/~jenrile/metadatamap/seeingstandards_glossary_pamphlet.pdf>
5. **EXIF - Exchangeable Image File Format** <http://en.wikipedia.org/wiki/Exchangeable_image_file_format> is a specification for the image file format used by digital cameras and scanners. The specification uses

the existing JPEG, TIFF Rev. 6.0, and RIFF WAV file formats, with the addition of specific metadata tags. It is not supported in JPEG 2000, PNG, or GIF.

6. **FGDC CSDGM (14)** Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata <http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/index_html> The standard commonly referred to as FGDC (although FGDC is the maintenance agency, and CSDGM is the actual element set) is a large and early metadata standard for geospatial information created by agencies of the US federal government. The FGDC web site describes the scope of this standard as to allow users to “determine the availability of a set of geospatial data, to determine the fitness [of] the set of geospatial data for an intended use, to determine the means of accessing the set of geospatial data, and to successfully transfer the set of geospatial data.” The current production version of FGDC CSDGM is 2.0, from 1998. Since this time, an international standard for geospatial information (ISO 19115) has emerged. Plans have been announced to create a US national geospatial metadata standard as a profile of ISO 19115, and to create version 3.0 of CSDGM as an implementation of that. This work has not yet been finalized from: Glossary of Metadata standards by Jenn Riley, Devin Becker, Indiana University Libraries, Copyright 2009-2010 Jenn Riley <http://www.dlib.indiana.edu/~jenrile/metadatamap/seeingstandards_glossary_pamp_hlet.pdf>
7. **GCMD DIF (5)** - Global Change Master Directory Interchange Format (DIF) <<http://gcmd.nasa.gov/User/difguide/>> A DIF consists of a collection of fields which detail specific information about data.
8. **ISO 14xxx Family of International Standards** on environmental management <http://www.iso.org/iso/iso_14000_essentials> - The ISO 14000 family addresses various aspects of environmental management
9. **ISO 19xxx Family (8)** <<http://www.isotc211.org/>> - A structured set of standards (created by committee for the ISO Technical Committee for Geographic Information/Geomatics) for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth. These standards may specify, for geographic information, methods, tools and services for data management (including definition and description), acquiring, processing, analyzing, accessing, presenting and transferring such data in digital/electronic form between different users, systems and locations. The work shall link to appropriate standards for information technology and data where possible, and provide a framework for the development of sector-specific applications using geographic data.
10. **MARC - Machine Readable Cataloging** - <<http://www.loc.gov/marc/>> - MARC was first developed in the late 1960s at the Library of Congress, and represented the first major attempt to encode bibliographic data in machine-readable form. MARC uses a mixture of fixed and variable fields to record information. The variable fields are themselves a mixture of coded and textual data. The MARC format is defined in ISO2709, which prescribes numeric field names that contain alphanumeric subfields. The MARC format in use in the US is known as MARC21. UNI- MARC is a variant common in Europe. While there are five formats in the MARC21 suite, the Bibliographic and Authority formats are the most commonly used. from: Glossary of

Metadata standards by Jenn Riley, Devin Becker, Indiana University Libraries,
Copyright 2009-2010 Jenn Riley
<http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards_glossary_pamphlet.pdf>

11. **MODS** - Metadata Object Description Schema -

<<http://www.loc.gov/standards/mods/>> - MODS was developed by the Library of Congress Network Development and MARC Standards Office as a MARC-compatible metadata format expressed in XML and using language-based element names. MODS takes a similar approach to resource description as MARC, with some rearranging, removing, and adding of data elements. MODS is frequently used as a descriptive metadata structure standard inside METS metadata wrappers for storage or exchange of digital objects. from: Glossary of Metadata standards by Jenn Riley, Devin Becker, Indiana University Libraries, Copyright 2009-2010 Jenn Riley
<http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards_glossary_pamphlet.pdf>

Appendix E

SURVEY RESULTS: TOOLS

NOTE: The survey seemed to show that metadata writers manage and produce metadata using existing tools and by developing their own tools using various databases, programming and markup languages. Some tools allow you to manipulate metadata standards, and other tools allow you to manipulate the controlled vocabularies that populate those standards -- but they are all tools for generating useful metadata.

1. **ArcCatalog metadata editor**, ESRI (3) form-based metadata entry interface designed to facilitate creation of FGDC CSDGM metadata within ArcGIS. At ArcGIS 10 changes to standard-independent internal structure
2. **CNS, USGS** <<http://geology.usgs.gov/tools/metadata/>> software for correcting some types of formatting errors that in years past were commonly encountered in FGDC CSDGM
3. **Environmental Metadata Editor - EME, EPA**
<<http://www.epa.gov/geospatial/eme.html>> The EPA Metadata Editor (EME) was developed to simplify and standardize geospatial metadata development across the Agency. It allows users to create and edit geospatial metadata records that meet the EPA Geospatial Metadata Technical Specification and Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) requirements.
4. **FGDC Geospatial Metadata Tools**- a step-by-step process to determine which tool is best for an implementation. <<http://www.fgdc.gov/metadata/geospatial-metadata-tools/?searchterm=Metadata%20Editors>>
5. **GeoNetwork** (2) <<http://geonetwork-opensource.org/>> GeoNetwork is a catalog application to manage spatially referenced resources. It provides powerful metadata editing and search functions as well as an embedded interactive web map viewer. It is currently used in numerous Spatial Data Infrastructure initiatives across the world.
6. **Geoparser** [A retired project designed to parse various types of documents and match them with various geographic thesauri. See "Unlock Service."
7. **Mercury Distributed Metadata Management, Data Discovery and Access System**
<<http://mercury.ornl.gov/>>
8. **MERMAid - Metadata Enterprise Resource Management Aid**
<<http://www.ncddc.noaa.gov/activities/mermaid>> MERMAid is a web-based tool developed at NCDDC that allows coastal data providers to create and manage metadata.
9. **Metascribe, NOAA/CSC**
<<http://www.csc.noaa.gov/metadata/metascribe/http://www.csc.noaa.gov/metadata/m>

[etascibe/](#)> The NOAA Coastal Services Center's MetaScribe tool is designed to reduce significantly the labor required to produce metadata compliant with the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM). This software tool takes advantage of the fact that, within a collection of records for a given data type, the records are generally very similar in content, with only a few fields or phrases changing from one record to the next. MetaScribe delivers the greatest savings when used to create a collection of similar records. In fact, this tool is not appropriate for the creation of multiple metadata records with little redundant content. MetaScribe is template driven (see example template). The user must create a metadata template, which is uploaded into MetaScribe. Building a template is not a trivial task; however, once a template is created for a given data type, the user can create multiple records quickly and easily. MetaScribe uses two proofing tools to check the records it creates: cns (Chew and Spit) and mp (Metadata Parser). These programs, created by the U.S. Geological Survey (USGS), will report any errors within a record and require that those errors be fixed before a final metadata record is produced. This illustrates the importance of creating a proper template that will produce error-free records.

10. **Metavist**, USFS (3) <<http://ncrs.fs.fed.us/pubs/products/metavist>> USDA Forest Service software tool for creating metadata compliant with two of the Federal Geographic Data Committee (FGDC) metadata standards: "FGDC Content Standard for Digital Geospatial Metadata" (FGDC 1998) and "FGDC Biological Data Profile of the Content Standard for Digital Geospatial Metadata"
11. **MODS Service - Metadata Object and Description Schema (MODS)**, a MARC-compatible XML schema for encoding descriptive data. The Library of Congress' Network Development and MARC Standards Office developed MODS (www.loc.gov/standards/mods/) in consultation with interested experts to satisfy the expressed need for an abbreviated XML version of MARC 21.
12. **MP**, USGS (4) <<http://geology.usgs.gov/tools/metadata/>> allows metadata editors to evaluate how well their metadata records conform to the structure specified by the FGDC CSDGM metadata standard.
13. **MP in Arc** (ESRI uses the USGS software "mp" in some cases to ingest and possibly to export FGDC CSDGM metadata)
14. **National Park Service metadata tools, NPS**
<http://science.nature.nps.gov/im/units/mwr/gis/metadata/metadata_tools.htm>
Specialized extensions to ArcMap and ArcCatalog intended to assist ArcGIS users to create and manage FGDC CSDGM metadata.
15. **OPeNDAP** is a framework that simplifies all aspects of scientific data networking. OPeNDAP provides software, which makes local data accessible to remote locations regardless of local storage format. OPeNDAP also provides tools for transforming existing applications into OPeNDAP clients (i.e., enabling them to remotely access OPeNDAP served data). <<http://www.opendap.org>>
16. **Protégé** <<http://protege.stanford.edu/>> a free, open source ontology editor and knowledge base framework.

17. **SMMS metadata editing software** for FGDC-CSDGM now owned by Intergraph and incorporated into their GeoMedia product
18. **TKME**, USGS (3) <<http://geology.usgs.gov/tools/metadatas/>> a general, extensible editor for FGDC CSDGM metadata
19. **Unlock Service** <<http://unlock.edina.ac.uk/>>Middleware designed for use by a web site or other application to search for place information found in various geographic information services
20. **USGIN Metadata Wizard** - Generic tool for creating useful minimum geospatial metadata and exporting metadata in multiple metadata standards and formats (beta release). <<http://mw.usgin.org/>>
21. **Web Accessible Folders of ISO 19115 XML files.** A Web Accessible Folder (WAF) is an HTTP accessible directory of files, typically metadata files in XML format in which all files and their time-stamps are visible to a web browser or client. Crawlers are able to parse the file listings and date-time stamps and provide a search interface on these documents.
<http://seabass.ieee.org/groups/geoss/index.php?option=com_sir_200&Itemid=157&ID=183>

22. Generic Tools Used with Metadata

- Common File Formats: Excel, Generic Text Editors, Word, PDF, E-mail
- Database Programs: MySQL, Oracle
- Scripting Languages: Perl, PHP
- Transfer Protocols: FTP, HTTP
- Mark-up Languages and related tools:
 - i. XML (2) <<http://www.w3.org/XML/>>interchange format for metadata and controlled vocabularies
 - ii. XML Notepad <<http://xmlnotepad.codeplex.com/>> a general XML document editor for browsing and editing XML documents.
 - iii. XPath a standard expression syntax used with software for finding specific information in an XML document

Appendix F

SURVEY RESULTS: VOCABULARIES

1. **ADL Feature Type Thesaurus (2)**
<<http://www.alexandria.ucsb.edu/~lhill/FeatureTypes/>> a formal thesaurus of types of named geographic features developed by the Alexandria Digital Library project
2. **American Geologic Institute (AGI) Glossary of Geology and AGI GEOREF -**
<<http://www.agiweb.org/pubs/pubdetail.html?item=300154>>
3. **Biological and Chemical Oceanography Data Management Office (BCO-DMO) common term lists -** <<http://bcodmo.org/>> (not publicly available)
4. **CF (Climate and Forecast Metadata Standard Names)**
<<http://cf-pcmdi.llnl.gov/documents/cf-standard-names>> CF is a controlled vocabulary for climate and forecast parameters.
5. **CIESIN list of Socio-Economic Keywords**
<http://sedac.ciesin.columbia.edu/metadata/vocab/vocab_intro.html>, Center for International Earth Science Information Network (CIESIN) has developed a list of socio-economic keywords and provides guidance on how to add them to FGDC CSDGM metadata.
6. **GCMD Science Keywords (8)**
<http://gcmd.nasa.gov/Resources/valids/archives/keyword_list.html>. GCMD keyword vocabularies have been developed and are constantly being refined and expanded. These vocabularies are also being in other applications within the science community. (from <<http://gcmd.nasa.gov/Aboutus/index.html>>).
7. **GEBCO Gazetteer of Undersea Feature Names**
<http://www.gebco.net/data_and_products/undersea_feature_names/> Undersea feature name in the form of a digital gazetteer.
8. **General Multilingual Environmental Thesaurus (GEMET)**
<<http://www.eionet.europa.eu/gemet>> - from the European Environment Agency. The GEMET vocabulary ensures validated indexing, cataloguing and retrieval within environmental information services as well as harmonized translations in the multilingual European network. It provides a common terminology system of controlled keywords which represent broad environmental concepts.
9. **GeoNames** <<http://www.geonames.org/>> The GeoNames geographical database covers all countries and contains over eight million place names that are available for download free of charge.
10. **GEOSS (Global Earth Observation System of Systems) Standards and Interoperability Registry,** <<http://seabass.ieee.org/groups/geoss/>>

11. **ISO 19115 Topic Categories or Keywords** (2) an extremely general, small (19 terms) set of high-level categories for data. See http://gcmd.nasa.gov/User/difguide/iso_topics.html or <http://marinemetadata.org/references/isotopics> for the list of categories. The FGDC Metadata Quick Guide <http://www.fgdc.gov/metadata/documents/MetadataQuickGuide.pdf> provides guidance for using the ISO 19115 Topic Categories in the Theme Keyword element (1.6.1.1) of CSDGM.
12. **Local lists of terms** (3) Oftentimes repositories keep and maintain local lists of terms used in their holdings. These terms can then be mapped to more global vocabularies while maintaining the integrity and meaning of the local terms intact.
13. **Marine Realms Information Bank/Coastal Change Hazards Digital Library controlled vocabularies** (4) <http://mrrib.usgs.gov/doc/facets.html>
14. **MMI Device Ontology** (4) <http://marinemetadata.org/community/teams/ontdevices> an ontology aimed at classifying sensor/device types, and possibly extending into more detailed sensor terminology. Interested members of the marine science and data community are developing this project.
15. **NBII National Biological Information Infrastructure** <http://www.nbii.gov/> (3), a broad, collaborative program to provide increased access to data and information on the nation's biological resources. The NBII links diverse, high-quality biological databases, information products, and analytical tools maintained by NBII partners and other contributors in government agencies, academic institutions, non-government organizations, and private industry.
 - a. **CSA/NBII Biocomplexity Thesaurus** <http://nbii-thesaurus.ornl.gov/thesaurus/> The Biocomplexity Thesaurus displays terminologies and term relationships in the fields of biology, ecology, environmental sciences, and sustainability.
 - b. **CERES NBII Thesaurus** <http://ceres.ca.gov/thesaurus/>: A common, integrated controlled vocabulary for the description, discovery, and exchange of environmental information
 - c. **Integrated taxonomic Information System (ITIS)** <http://www.itis.gov/> authoritative taxonomic information on plants, animals, fungi, and microbes of North America and the world.
16. **Q2O for QA/QC** <http://q2o.whoi.edu/> - A project working to enable quality assurance / quality control standards for marine-related data into existing frameworks such as Open Geospatial Consortium (OGC) and Sensor Web Enablement (SWE).
17. **SeaVox** - https://www.bodc.ac.uk/data/codes_and_formats/seavox/ - a simple and effective governance process subscribed to by a community of people working for repositories and other organizations that employ a defined set of SeaDataNet and BODC vocabularies (SeaDataNet is mentioned in section on Crosswalks).

18. **U.S. Board on Geographic Names (BGN)** - "The U.S. Board on Geographic Names is a Federal body created in 1890 and established in its present form by Public Law in 1947 to maintain uniform geographic name usage throughout the Federal Government. The Board comprises representatives of Federal agencies concerned with geographic information, population, ecology, and management of public lands. Sharing its responsibilities with the Secretary of the Interior, the Board promulgates official geographic feature names with locative attributes as well as principles, policies, and procedures governing the use of domestic names, foreign names, Antarctic names, and undersea feature names." <<http://geonames.usgs.gov/>>
- a. **Domestic Names** Domestic Names - Geographic Names Information System (GNIS) - The GNIS is the Federal and national standard for geographic nomenclature. The GNIS is the official repository of domestic geographic names data, the official vehicle for geographic names use by all departments of the Federal Government, and the source for applying geographic names to Federal electronic and printed products.
<<http://geonames.usgs.gov/domestic/>> See FGDC CSDGM metadata for GNIS <<http://geonames.usgs.gov/domestic/metadata.htm>>
 - b. **Antarctic Names** - <<http://geonames.usgs.gov/antarctic/>>
 - c. **Foreign Names** - Information about foreign geographic feature names can be obtained from the GEOnet Names Server (GNS), developed and maintained by the National Geospatial-Intelligence Agency (NGA). The GNS database is the official repository of foreign place-name decisions approved by the U.S. Board on Geographic Names. <<http://earth-info.nga.mil/gns/html/>>
 - d. **Undersea Features** - Information about undersea feature names and the Advisory Committee on Undersea Features of the U.S. Board on Geographic Names can be obtained from the GEOnet Names Server (GNS), developed and maintained by the National Geospatial-Intelligence Agency (NGA).
<<http://geonames.usgs.gov/undersea/>>
19. **USGS Geologic Names Lexicon** <<http://ngmdb.usgs.gov/Geolex/>> database describing historical usage of geologic formation names
20. **USGS Thesaurus** (5) <<http://www.usgs.gov/science/about/>> a broad, shallow formal thesaurus intended to help people outside USGS find scientific information resources without having to know the organizational structure of USGS itself
21. **VLIMAR (VLIZ Marine Gazetteer)**: VLIZ developed a standard, relational list of geographic names, coupled with information and maps of the geographic location of these features. The purpose of the gazetteer is to improve access and clarity of the different geographic, mainly marine names such as seas, sandbanks, ridges, bays or even standard sampling stations used in marine research. The geographic cover is global, however the gazetteer is focused on the Belgian Continental Shelf, the Scheldt Estuary and the Southern Bight of the North Sea.
<<http://www.vliz.be/vmdcdata/vlimar/index.php>>
22. **VSTO** <<http://www.vsto.org/>> The Virtual Solar Terrestrial Observatory vocabulary and ontology (VSTO) provides a set of terms and relationships between those terms. A

graphical representation of version 1 of this information can be found here :
<http://tw.rpi.edu/proj/portal.wiki/images/1/1e/VSTO_1.0_Ontology_Diagram.pdf>

Version 2 is available at <<http://escience.rpi.edu/ontology/vsto/2/0/>>

Appendix G

DESCRIPTIONS OF VOCABULARY-RELATED RESOURCES IN OTHER PROJECTS

1. **ANDS Vocabulary (workshop)**
<<http://community.ands.org.au/viewforum.php?f=101>>
2. **CGI (Commission for the Management and Application of Geoscience Information)**
<<http://www.cgi-iugs.org/>>
3. **EPA System of Registries - Vocabularies program**
4. **European Atlas of the Seas**
<http://ec.europa.eu/maritimeaffairs/atlas/index_en.htm>
With the growing importance of maritime affairs on the political agenda of EU Member States, a new political geography has been created. Supported by the Spanish Ministry of Education and Science, Juan Luis Suárez de Vivero and Juan Carlos Rodríguez Mateos have developed an Atlas of the European Seas and Oceans to record this new geography in maps and graphs. In addition to territory-related issues, the Atlas includes a summary of all the main marine uses and activities from a European perspective. The Commission has published an English version of the Atlas with the kind permission of the authors and in line with the strategy proposed in An Integrated Maritime Policy for European Union (The Blue Book) (October 2007). <http://ec.europa.eu/maritimeaffairs/atlas/index_en.htm>
5. **International Coastal Atlas Network (ICAN)**
ICAN is a group of organizations that have been meeting over the past four years to scope and implement data interoperability approaches to coastal web atlases (CWAs). The **mission/strategic aim** of ICAN is to share experiences and to find common solutions to CWA development (e.g., user and developer guides, handbooks and articles on best practices, information on standards and web services, expertise and technical support directories, education, outreach, funding opportunities, and so on), while ensuring maximum relevance and added value for the end users.
<<http://www.icoastalatlans.net>>
6. **Long Term Ecological Research Network (LTER)** <<http://www.lternet.edu/>>
LTER is an ongoing ecology collaboration that has developed standards and resources to facilitate the stewardship, curation, and sharing of data. (For more information see article: "Enriching the notion of data curation in E-science: data managing and information infrastructuring in the Long Term Ecological Research (LTER) Network.")
<<http://interoperability.ucsd.edu/docs/06KarastiBakerJCSCW.pdf>>
7. **Marine Metadata Initiative (MMI)**
It was suggested that that this "small" cmisp effort might use MMI as a base.
8. **National Climate Assessment**
Also going through a vocabulary assessment process.

9. **National Geoinformatics Community** <<http://ngc.usgin.org/>> is a community of practice formed under the auspices of the U.S. Geoscience Information Network (GIN) <<http://usgin.org/>>. The GIN is a system of state and federal geological survey online data providers focused on developing standardized web services, interchange formats, and tools to promote the discovery and use of geoscientific information. In addition to supporting these larger GIN goals, the NGC is also looking at how to identify, select, manage, and apply suitable vocabularies, for data discovery, evaluation, and integration in earth systems science.

10. **West Coast Coastal Atlas Group**

This is regional subgroup of ICAN that has been looking at cross-border data sharing and region-wide data needs. An April 2009 report includes a very nice workshop summary document with good summary of region-wide data needs <http://dusk.geo.orst.edu/ICAN_EEA/WestCoast/West_Coast_Atlasses_Workshop.pdf>. The group continues to meet via Webinars for information sharing and coordination. <<http://www.icoastalatlantlas.net/westcoast>>

Appendix H

CRITERIA FOR EVALUATING INTRINSIC ASPECTS OF VOCABULARIES

1. **Clarity:** An ontology should effectively communicate the intended meaning of defined terms. Formalism is a means to this end. When a definition can be stated in logical axioms, it should be. Where possible, a complete definition (a predicate defined by necessary and sufficient conditions) is preferred over a partial definition (defined by only necessary or sufficient conditions). All definitions should be documented with natural language. VRL.
2. **Objectivity:** Definitions should be objective. While the motivation for defining a concept might arise from social situations or computational requirements, the definition should be independent of social or computational context.
3. **Coherence:** An ontology should be coherent: that is, it should sanction inferences that are consistent with the definitions. At the least, the defining axioms should be logically consistent. Coherence should also apply to the concepts that are defined informally, such as those described in natural language documentation and examples. If a sentence that can be inferred from the axioms contradicts a definition or example given informally, then the ontology is incoherent. VRL.
4. **Extendibility:** An ontology should be designed to anticipate the uses of the shared vocabulary. It should offer a conceptual foundation for a range of anticipated tasks, and the representation should be crafted so that one can extend and specialize the ontology monotonically. In other words, one should be able to define new terms for special uses based on the existing vocabulary, in a way that does not require the revision of the existing definitions. VRL.
5. **Minimal encoding bias:** The conceptualization should be specified at the knowledge level without depending on a particular symbol-level encoding. An encoding bias results when representation choices are made purely for the convenience of notation or implementation. Encoding bias should be minimized, because knowledge sharing agents may be implemented in different representation systems and styles of representation. VRL.
6. **Minimal ontological commitment:** An ontology should require the minimal ontological commitment sufficient to support the intended knowledge sharing activities. An ontology should make as few claims as possible about the world being modeled, allowing the parties committed to the ontology freedom to specialize and instantiate the ontology as needed. Since ontological commitment is based on consistent use of vocabulary, it can be minimized by specifying the weakest theory (allowing the most models) and defining only those terms that are essential to the communication of knowledge consistent with that theory.

7. **Contextual Relevance:** can be assigned by number and degree to which key concepts/ terms are defined in the CMSP context. VRL
8. **Maturity:** in relation to CMSP need and perhaps in comparison to other vocabularies, number of years, number of users, applications, revision level, community convention or standard, national or international standard, etc. VRL
9. **Original Intend Use:** degree to which CMSP use is aligned or not with original intended use, an estimate of how 'ready' the vocabulary is and some indication of degree of alignment (text)
10. **Fitness for Use:** estimate of current fitness for purpose for CMSP (before modification or extension), especially related to accuracy of vocabulary and minimal alteration, extension required. VRL (comments in text)
11. **VRL: Vocabulary Readiness Level**
Applying the principles to derive a VRL:
 - VRL 1-3** - assessment evidence - little to none in application or service
 - VRL 4-6** - assessment evidence in demonstrated application
 - VRL 7-9** - widely available and used in application or service

First 5 criteria are from Gruber, Toward Principles for the Design of Ontologies Used for Knowledge Sharing (1993) <<http://tomgruber.org/writing/onto-design.pdf>>

Appendix I

ACRONYMS

BCO-DMO	Biological and Chemical Oceanography Data Management Office
CIESIN	Center for International Earth Science Information Network (CIESIN)
CMSP	Coastal and Marine Spatial Planning
CSDGM	Content Standard for Digital Geospatial Metadata
FGDC	Federal Geographic Data Committee
GCMD	NASA's Global Change Master Directory
HTTP	Hypertext Transfer Protocol
IOOS	Integrated Ocean Observing System (led by NOAA)
ISO	International Organization for Standardization
ILTER	Long Term Ecological Research (ILTER) Network
MERMAid	Metadata Enterprise Resource Management Aid NOAA/NCDDC
MMI	Marine Metadata Initiative
MODS	Metadata Object Description Schema
NBII	National Biological Information Infrastructure
NOAA	National Oceanic and Atmospheric Administration
NOC NIMS	National Ocean Council National Information Management System (NOC NIMS)
R2R	Rolling Deck to Repository: NSF-funded project aimed at developing a comprehensive fleet-wide management of underway data to ensure preservation of and access to our national oceanographic research data resources from UNOLS vessels.
RDF	Resource Description Framework (http://www.w3.org/RDF/)
RPI	Rensselaer Polytechnic Institute
URI	Uniform Resource Indicator (http://en.wikipedia.org/wiki/Uniform_Resource_Identifier)
USGS	United States Geological Survey
VRL	Vocabulary Readiness Level
WHOI	Woods Hole Oceanographic Institution
XML	Extensible Markup Language (http://www.w3.org/TR/2000/REC-xml-20001006)

Appendix J

Agenda

Wednesday, December 1

- 12:00 Opening Lunch
- 1:00 Introductions and Review of Meeting: Andy Maffei, WHOI
- 1:30 “Overview of the National Ocean Policy, the Coastal and Marine Spatial Planning Framework and its Data Management Requirements”: Betsy Nicholson and Tony Lavoie, NOAA Coastal Services Center
- 2:15 Coffee Break
- 2:45 “Interoperable Data Catalogs and Discovery Vocabularies”: Peter Fox
- 3:45 Overview of Key Resources and Issues Identified by Participants
- 4:15 Review of Breakout Group Topics
- 5:00 Break
- 6:00 Reception and Dinner in Clark 507

Thursday, December 2

- 8:30 Breakfast
- 9:00 Workshop Outcomes: Discussion and Goal Setting
- 12:00 Lunch
- 1:30 Charge to Groups
- 2:00 Group Work Time
- 3:30 Report from Groups
- 5:00 Adjourn

Friday, December 3

- 8:30 Breakfast
- 9:00 Synthesis Panel and Discussion: Recommendations for shared vocabularies, crosswalks and ontologies
- 9:45 Coffee Break
- 10:00 Synthesis Panel and Discussion: Next steps and future communication/coordination strategies
- 11:00 “Knowledge Provenance”: Peter Fox, RPI
- 12:00 Adjourn