

## Metadata Survey Findings for Most Regional OOS

Contact: Stephanie Watson, Steph\_Watson@sbcglobal.net

[Direct link to Summary](#)

### Introduction

This document summarizes the results of two surveys on how regional and sub-regional ocean observing systems are addressing metadata needs: a) a data management survey distributed by the NOAA Coastal Services Center to the PIs of projects funded by NOAA Coastal Observation Technology System (COTS) program; and b) a brief metadata survey sent to the regional ocean observing systems in IOOS (with the exception of the Caribbean due to their very nascent nature). The summary of the survey results, along with the initial IOOS Data Management and Communications (DMAC) guidelines (1) and progress from the Marine Metadata Interoperability (MMI) initiative, will be used to help inform a metadata best practices document for the Gulf of Mexico Coastal Ocean Observing System (GCOOS). Although funded by a contract to GCOOS through the Southeastern Universities Research Association (SURA), this information is also freely available to other regional ocean observing systems.

As a precursor to the November, 2004 NOAA COTS workshop, the NOAA Coastal Services Center distributed a survey to the COTS-funded PIs on their data management practices. The survey asked the data managers to provide a brief description of their project's data management methods, including metadata practices.

In July, 2005 a brief, specific metadata survey was sent via email to the data management experts of nine of the eleven regional ocean observing systems in the U.S. Integrated Ocean Observing System (IOOS). The survey included the following nine questions:

1. Are you using (a) particular metadata standard (or standards) to document your data sets (e.g., Federal Geographic Data Committee - FGDC, Directory Interchange Format - DIF, International Standards Organization (ISO) 19115/19139, Dublin Core, etc.)? If you are using a particular FGDC profile or extension, please indicate which.
1. Does the standard you are using include fields for syntactic metadata? (Syntactic metadata describes what the data \*looks\* like and how it is organized. Syntactic fields often include the unit of measurement, data type – such as int, float, etc.)
1. Does the standard you are using include fields to capture semantic metadata? (Semantic metadata describes what the data really means, such as long name, definition, etc.)
1. Do you use any specific tools to help develop and validate metadata, as well as submit it to a metadata catalog? If yes, please name them.
1. What format are your metadata records in (text, xml, html, etc...)? If you are using XML, please indicate the particular XML schema and stylesheet you are using.
1. How are you storing your metadata records (file directory, database, etc.)?
1. How are the metadata provided to users (header to actual data file, bundled with data file during download or other transfer, etc...)
1. Are you using a metadata catalog? If yes, which one? Is it connected to the National Spatial Data Infrastructure or Geospatial Onestop?

- I. If necessary, is it ok to follow up with you for additional information?

The following section outlines the responses from each regional ocean observing system (and some of their sub-regional systems, in particular) and concludes with summarized results for all the systems. Please note that: a) a summarized discussion of semantic and syntactic metadata will be included at the end of this section; b) due to the short timeline for this survey, some responses did not include information for all of the questions above; and c) it is still not clear to some of the regional (and sub-regional) OOS, which data will be part of their systems.

## Results

### ***Gulf of Mexico Coastal Ocean Observing System (GCOOS)***

#### Content standard

- Texas Automated Buoy System (TABS) (via NOAA CSC), the West Florida Shelf Coastal Ocean Monitoring and Prediction System (COMPS) – University of South Florida, and the Wave-Current-Surge Information System (WAVCIS) using FGDC
- NetCDF headers (for use with.opendap/LAS) with TABS

#### Tools

- COMPS is using Cast-Net (precursor to Metadoor). Also looking at other tool options, such as SMMS (a licensed tool developed Intergraph and being used by Florida Fish and Wildlife Research Institute) and MERMAid (developed by NOAA National Coastal Data Development Center)

#### Format

- COMPS is using XML

#### Metadata provided to users

- TABS offers downloadable ASCII files and charts of the data. Some metadata are provided as column headings in the ASCII file and as labels in the chart.

#### Storage

#### Catalog

### ***Mid Atlantic Coastal Ocean Observing Regional Association (MACOORA)***

#### Content standard

- Chesapeake Bay Observing System provides metadata for ADCP consistent with USGS practices (unclear what this means)

#### Tools

#### Format

#### Metadata provided to users

#### Storage

- database for the New York New Jersey Harbor Ocean Prediction System

#### Catalog

### ***Southeast Atlantic Ocean Observing System (SEACOOS)***

#### Content standard

- FGDC
- SensorML (not really a standard, but a convention for metadata concerning platforms and sensors)
- MarineXML

- O&M (for metadata concerning observations and measurements)
- OGC-ISO
- SEACOOS CDL NetCDF convention (standardizes metadata in headers)
- DIF

#### Tools

- 'Meta-Door' for FGDC record development and publishing
- 'mp' tool (metadata parser) provided by USGS for validating FGDC records

#### Format

- FGDC xml format
- Moving to support SensorML and MarineXML
- Also trying to promote use of ASCII column descriptive XML for use in parsing and translating

#### Storage

- both file and database (Postgres relational database, including the sub-regional systems)
- would like to move towards internal database representation which can be mapped or translated to multiple file (xml and otherwise) formats.

#### Metadata provided to users

- General keyword search metadata available via FGDC search
- Data-specific metadata is available via file or metadata header.

#### Catalog

- FGDC records published to FGDC clearinghouse catalog and harvested by geodata.gov
- Other
  - units declared as Unidata udunits

### ***Gulf of Maine Ocean Observing System (GoMOOS)***

#### Content standard

- GoMOOS buoys in FGDC
- GoMOOS uses ISO standard in MapServer applications
- Coastal Ocean Observation and Analysis (COOA) uses DIF

#### Tools

- GoMOOS uses mp for FGDC validation

#### Format

- Text
- HTML
- XML (for use with MapServer)
- SGML
- NetCDF (for GoMOOS sensor metadata)

#### Storage

#### Metadata provided to users

#### Catalog

- COOA uses the Global Change Master Directory (GCMD)
- GoMOOS will use GeoConnections, GCMD, Geo-spatial Onestop

#### Other

- COOA using GCMD keywords

**Central and Northern California Ocean Observing System (CeNCOOS)  
(specifically, the Center for Integrated Marine Technologies or CIMT and the  
Center for Integrated Coastal Observation, Research and Education or CICORE)**

Content standard

- FGDC for CICORE, as well as CIMT ship data and SST satellite imagery
- COARDS compliant NetCDF for CIMT moorings
- Goddard DAAC for satellite ocean color
- some SEACOOS CDL NetCDF for CICORE

Tools

- ArcCatalog metadata tools for CICORE, MBARI's Shore Side Data System for CIMT

Format

- text for CICORE and CIMT

Storage

- file directory for CICORE and CIMT

Metadata provided to users

- download in separate file for CICORE

Catalog

- MBARI's Shore Side Data System for CIMT

Other

- using keywords from GCMD, GLOBEC, and JGOFS

**Southern California Coastal Ocean Observing System (SCCOOS)**

Content Standard

- Coastal Data Information Program (CDIP) - OGC-ISO
- Since 1995, CDIP's wave metadata has conformed with USACE Wave Data Analysis Standard
- Metadata for CDIP, the SCCOOS Manual and Automated Shore Stations, and the Point La Jolla Mooring are compliant with FGDC

Tools

- CDIP, the SCCOOS Manual and Automated Shore Stations, and the Point La Jolla Mooring use mp for validation

Format

- CDIP uses HTML and XML

Storage

- CDIP, the SCCOOS Manual and Automated Shore Stations, and the Point La Jolla Mooring use a database

Metadata provided to users

- CDIP provides metadata via its web page. Metadata are separate from the data files
- For the SCCOOS Manual and Automated Shore Stations and the Point La Jolla Mooring, the metadata and data are stored together in a relational database. Both are then available via a data query webpage on the SCCOOS website. When queried, metadata are passed through the mp tool and served as either HTML- or XML-formatted data as requested by the user.

Catalog

- Neither CDIP, the SCCOOS Manual and Automated Shore Stations, nor the Point La Jolla Mooring are currently using a catalog, but might in the future

## **Alaska Ocean Observing System (AOOS)**

### Content standard

- May use Ecological Metadata Language (EML) with translators to FGDC (for the GEM subregional effort). EML has been chosen for this purpose because it is a superset of metadata protocols such as FGDC, Z39.50, and Dublin Core and provides a structure for advanced data set documentation
- FGDC
- SEACOOS CDL for NetCDF

### Tools

- no tools used

### Format

- text
- NetCDF
- database (goal to work with Geography Markup Language's XML Schema)

### Metadata provided to users:

- header to datafile
- separate file bundled with data file during download or other transfer
- user specified delimited file with header

### Storage

- files and Postgres database

### Catalog

- no

## **Great Lakes Ocean Observing System (GLOS)**

### Content standard

- reference to FGDC (with ISO 19139 harmonization in the future)

### Tools

- ArcCatalog for development

### Format

- XML metadata sans schema/stylesheets are the preferred format, being by far the most modifiable (with stylesheets) and viewable (being ASCII) option

### Data provided to users

- Inclusion of the metadata as a separate file bundled with each dataset

### Storage

- We have metadata records within both file directories and databases. Some metadata records are documented within a file attached to the data file, while other metadata records are stored as fields within a database. We are currently migrating our GIS files from local storage system to ESRI SDE and it has the capabilities to import metadata fields and relate to a specific GIS layer. In this way, the GIS file and the metadata record will be stored within a database.

### Catalog

- We are currently not using a metadata cataloging system. We made attempts to utilize the Z39.50 server and it turns out that it is not very effective when data and metadata are stored within a database. We currently have the newest version of ESRI ArcIMS, which also has the capabilities to implement a metadata catalog service. As we

begin to migrate our data from a file directory and into ESRI SDE, the metadata catalog service will be able to serve external users. Since the Geospatial Onestop also utilize the ESRI ArcIMS service, it will easily integrate with the NSDI.

### **Northwest Association of Networked Ocean Observing Systems (NANOOS)**

Content standard

- FGDC

Tools

Metadoor for FGDC metadata template creation., with USGS mp tool for validation. Perl scripts were developed based on the templates to automatically generate FGDC compliant records based on existing metadata in files and relational databases. Compliance is verified using the USGS mp tool

Format

- ASCII text
- HTML
- XML

Data provided to users

- The FGDC metadata records are accessed by users via links in web pages (see <http://www.ccalmr.ogi.edu/CORIE/data/publicarch/>). Links to the data are provided within each FGDC record using both the "Olinke\_Linkage" and "Network\_Resource\_Name" fields. Additionally, each data file has a header with non-compliant metadata.

Storage

- relational database
- file storage (directories)
- concurrent version systems repositories (CVS, <https://www.cvshome.org/>)
- Directories are for web access and metadata development, the CVS repository is for version control, and the relational databases are for applications. There are NANOOS participants using XML database stores for applications.

Catalog:

- No - we maintain our "node catalog" and have not connected to NSDI or Geospatial Onestop. We are investigating the FGDC Z39.50 lsite server and publishing to the FGDC Clearinghouse.

### **Pacific Islands Ocean Observing System (PacIOOS)**

Content standard

- FGDC
- None

Tools

The following were identified as specific tools to help develop and validate metadata, as well as submit it to a metadata catalog.

- ArcView® Metadata Collector v2.0 Extension
- MetaScribe
- Chew 'n Spit (CNS) metadata pre-parser
- Metadata Parser (MP)

- ESRI FGDC Metadata Editor
- ESRI's ArcCatalog
- CADD/GIS Technology Center's SDSFIE Geodatabase

#### Format

Metadata records exist in at least the following formats.

- Text Files
- XML , with FGDC CSDGM XML format and LAS XML noted as schema standards
- SGML
- HTML
- Netcdf (binary format with header info), EPIC specific netcdf, and Argo specific netcdf.
- MET (Corps of Engineers – text) metadata format.

#### Data provided to users

Metadata is bundled with data file during download/transfer, through http requests (to access html files), and through a registered Clearinghouse node.

#### Storage

- Locally in a file directory along with the actual dataset - as headers within the data sets (e.g., netcdf, grib) or are included with control files with the data (e.g. file directory, listed in XML catalogues for our servers), and
- On a registered FGDC Clearinghouse node.
- Efforts are also underway to store metadata in ArcIMS (Metadata Catalog) via ArcSDE.

#### Catalog

- In general, No. However one respondent reported that they will be building a metadata catalog in the future using ArcSDE & ArcIMS. Currently, they are using ArcCatalog. Files are uploaded to NSDI via an ftp, but this is done very irregularly.

### ***Southeastern Universities Research Association's Coastal Ocean Observing Program (SCOOP)***

#### Content standard

- FGDC
- COARDS Climate Forecast convention for NetCDF files
- OpENDAP header files

#### Tools:

#### Format:

#### Data provided to users

#### Storage:

#### Catalog:

#### Other

- use of controlled keywords, descriptive file naming conventions.
- SCOOP will also assist partners in developing Earth Science Markup Language (ESML) structural metadata descriptions of selected SCOOP data sets, and field an ESML registry for SCOOP in order to facilitate data interchange.

### ***Pacific Coast Ocean Observing System (PaCOOS)***

#### Content Standard

- FGDC

#### Tools

- Fisheries Information Systems (FIS) is working on a couple tools to search and catalog metadata. The Integrated Taxonomic Information System

(ITIS) goal is to create an easily accessible database with reliable information on species names and their hierarchical classification. A second effort, the FIS InPort Metadata Catalog, should give NOAA Fisheries and its state and regional partners the capability to share essential information about fisheries-dependent data.

#### Format

- Metadata records are slowly moving toward XML standard. Some of them are still in .doc format or .txt

#### Data provided to users

#### Storage

- Databases
- File directories

#### Catalog

- Not at this time. The PaCOOS semantic metadata effort goal is to create a data dictionary with web services to search metadata and datasets.

### **Summary**

#### Metadata standard

FGDC is the most common metadata standard used, followed by ISO 19115/19139 and DIF. Other less common metadata “standards” being used include: SensorML, MarineXML, SEACOOS NetCDF convention, COARDS CF NetCDF convention, and DODS/OpENDAP headers.

#### Syntactic metadata

Syntactic metadata describes what the data \*looks\* like and how it is organized. Syntactic metadata fields often include the unit of measurement, data type – such as int, float,, etc. Syntactic metadata are addressed by most of these “standards”.

#### Semantic metadata

Semantic metadata describes what the data really means, such as long name, definition, etc. Semantic metadata are addressed to some extent by these standards, but not in such a way to allow for sufficient interoperability. The progress of the MMI should help address this need.

#### Tools

Meta-Door is the most common tool used for metadata development and publishing, followed by ArcCatalog, Cast-Net, ArcView® Metadata Collector v2.0 Extension, ESRI FGDC Metadata Editor, MetaScribe, CADD/GIS Technology Center’s SDSFIE Geodatabase.

USGS’ mp tool is the most common tool for validation of metadata records to the FGDC standard. Some also use the Chew ‘n Spit (CNS) metadata pre-parser to prepare records for mp that cannot directly be parsed by mp.

#### Format

XML (without the use of a particular stylesheet) is the most common format for metadata records, followed by HTML, ASCII text, NetCDF, SGML, relational database, and MET.

## Metadata provided to users

Metadata are typically provided to users (primarily through http requests) in a file separate from the data file. The metadata may be bundled with the data file, but may not be. Metadata are less commonly provided in a data file header or can be accessed from a link embedded in the data file.

## Storage

Metadata records are most often stored in relational databases, followed by file directories, FGDC clearinghouse nodes, or concurrent version system repositories.

## Catalog publication

Most regional (and sub-regional) OOS are not publishing their metadata to a publicly-accessible metadata catalog, which would allow for data discovery. Those that do tend to publish to FGDC clearinghouses and the Geospatial Onestop.

### *Acknowledgements:*

Many thanks to the following for their feedback: Larry Atkinson, Antonio Baptista, Jim Boyd, Rob Cermak, Jeremy Cothran, Toby Garfield, Roger Gauthier, Jerome King, Allen Macklin, John Marra, Jan Newton, Tom Shyka, Julie Thomas, Paul Turner, and Jerry Wanetick, Funding provided by GCOOS, through a grant from the SURA Coastal Ocean Observing Program (SCOOP).

### (I) Initial DMAC Metadata Guidance:

- a. Guidance applies to both present data sources, and legacy data holdings and inventories.
- b. All metadata should be created in an XML-schema with an XML style sheet
- c. Metadata should be created that are compliant with the Federal Geographic Data Committee (FGDC). If FGDC extensions are not available for the specific type of data in question, use an alternative, community-accepted standard and document standard used.
- d. Data providers should be alert to their metadata needs, and should identify gaps so that their needs can be addressed through the Steering Team.
- e. Submit metadata to one of the National Spatial Data Infrastructure nodes where they will be widely available to the community.
- f. Metadata providers should document the data dictionary used.
- g. There are other substantial efforts already in place to address IOOS marine metadata needs from which the DMAC Metadata Expert Team will be drawing, including: the FGDC Subcommittee on Marine Metadata, the National Science Foundation-sponsored Marine Metadata Interoperability Initiative, the collaborative Office of Naval Research-NOAA Coastal Observation Technology System Working Groups, and the Southeastern Universities Research Association.
- h. The Steering Team recommends that an IOOS DMAC Expert Team on Metadata be formed, and that data providers participate in this team to ensure that their requirements for all types of data are met.