

Gulf of Maine Ocean Data Partnership

Technical Guidance for Data Sharing

Developed by the Technical Committee

Version 1.2

November 2007

Table of Contents

Purpose of Document	3
Obtaining Help – Partnership Assistance	3
Desired Datasets – Type and Geographic Scope of Data Collections	3
Data Assurance and Authority	3
Recommended Vocabulary (Standardized Names)	4
Species Names	4
Parameter Names	4
Gazetteer Names	5
Data Discovery – Constructing and Registering Metadata	5
Metadata Creation and Editing Tools	7
Data Transport and Access	8
Interoperability	9

Purpose of Document

The purpose of this document is to provide “nuts and bolts” guidance and specific recommendations in key areas needed for successful data sharing and interoperability. This document will be updated as experiences and emerging technologies dictate. For more information on how particular guidance was developed or modified over time, please see the annual technical committee guidance reports which are available on the Gulf of Maine Ocean Data Partnership’s (Partnership’s) web site: www.gomodp.org. Please check the version number on the front cover of this document to determine if you have the latest guidance.

Obtaining Help – Partnership Assistance

To obtain assistance in meeting data sharing criteria and recommendations, please contact:

- the Partnership’s host, the Gulf of Maine Ocean Observing System (GoMOOS), via Tom Shyka at tom@gomoos.org or www.gomoos.org; or
- another partner who has expertise in the area. To look for help from a partner with expertise in a particular area, use the Partnership’s online database and search tool located at <http://gomodp.who.edu/>. The database is keyword searchable, and will return a list of names and contact information for partners with expertise in the area you are looking (keywords such as: http, FDCG, java, etc).

Desired Datasets – Type and Geographic Scope of Data Collections

Initially, the Partnership wants datasets that:

- are multi-year or relatively stable datasets (such as one-time eel grass mapping)
- contain a surface water/ocean (not groundwater) component
- are within the Gulf of Maine watershed area. (However, important datasets for areas outside this watershed, such as Nova Scotia and south of Cape Cod, will not be excluded.)

Partners with a dataset that covers this watershed area and more are encouraged to share the entire dataset as opposed to filtering out data for this region alone.

Data Assurance and Authority

Each partner is ultimately responsible for quality assurance and quality control of data they share. Partners will be expected to adhere to the following:

- 1) Each partner will notify the Partnership of any issues related to ownership of data or data products that they contribute. The Partnership prefers datasets where sufficient authority exists to share the data. To determine if a dataset should be shared/contributed and to properly document ownership, the partner should consider the following questions:
 - Are you the primary creator/collector of the dataset or is there shared ownership?
 - If you are not the primary creator/collector, do you have a formal distribution agreement with the data provider?
 - Who else has the data or is serving them?
 - Who will correct errors in the event any are discovered?
 - Who should the data user contact prior to using the data (if applicable)?

- 2) Partners that enhance or create new products from external (not originally their own) data will indicate the original data source in metadata documentation and registration.
- 3) Partners will contribute and maintain the best available representation of their data.
- 4) Each partner will be responsible for remedying any problems within a dataset in a timely manner.
- 5) Partners are encouraged to properly preserve/archive their data for perpetuity. Proper data security includes data backup offsite to avoid loss in the case of a local event.
- 6) If a partner cannot or will no longer serve a dataset of interest to the Partnership, the partner will work with the Partnership to ensure the data are secured for future availability.
- 7) Partners will report problems encountered with other data sources served by the partnership. As data “users”, they will provide comments to the individuals who served the data.
- 8) If a partner can not fulfill any of these responsibilities, the partner will work with the Partnership to develop an alternative solution.

Recommended Vocabulary (Standardized Names)

Partners are encouraged to either use a standardized vocabulary within their data systems or create a crosswalk from their non-standardized vocabulary to a standardized one.

Species Names

- 1) For those partners not yet committed to a source of standard species names, the Integrated Taxonomic Information System (ITIS - <http://www.itis.usda.gov/>) is recommended.
- 2) For partners using other standard sources of species names, the partnership will initially look to the Universal and Biological Indexer and Organizer project's (UBIO) Taxonomic Name Server (see <http://www.ubio.org/SOAPbrowser/>) to provide the crosswalk functionality we require in the form of an automated service, at least on a limited basis.

Parameter Names

- 1) Use one of the standard sources mentioned on the Marine Metadata Interoperability web site (<http://marinemetadata.org/vocabularies/refs/markup/>). A current list with the document date is displayed below:

[The British Oceanographic Data Centre Parameter Usage Vocabulary](#)

The BODC Parameter Usage Vocabulary may be used as a tool to label a measurement with information on what it is and how it was obtained or to describe parameters (phenomena in GML terminology) in 'use' metadata.

[U.S. Joint Global Ocean Flux Study Parameters](#)

List of parameter names, descriptions, and units according to multiple categories (e.g., sampling attributes, physical properties, chemistries)

[IOC GF3 parameter codes](#)

Listing of parameter codes, subsetted according to CTD data, moored current meter data, etc.

[CF - Climate and Forecast Standard Names](#)

Climate and forecast (CF) standard names is a parameter vocabulary that is intended for use with climate and forecast data, for atmosphere, surface and ocean, and was designed with model-generated data particularly in mind.

[Data Dictionary from SEACOOS](#)

The current data dictionary with standard names used by SEACOOS. Includes standard names from many categories/disciplines, with extensive detail about each name.

[U.S. GLOBEC Thesaurus](#)

Alphabetical listing of field names with definitions and units associated with the U.S. GLOBEC Program.

Gazetteer Names

- 1) All datasets should contain (where possible) spatial coordinate data to allow for spatial queries.
- 2) It is also useful to have the data referenced to common geographic names for non-coordinate queries. Recommended vocabularies for geographic names are: the USGS Geographic Names Information System (<http://nhd.usgs.gov/gnis.html>) and the Geographical Names of Canada (http://geonames.nrcan.gc.ca/index_e.php).

Data Discovery – Creating and Registering Metadata

General guidelines for creating/registering metadata:

- 1) Each partner should register their own metadata since they are most knowledgeable and this is the most efficient way to complete the process.
- 2) Consult the guidance/help available on the portals or consult a GoMODP metadata resource person/organization to avoid errors and optimize metadata entry.
- 3) Peer and supervisor review of metadata is recommended to ensure accuracy and desired level of technical detail.
- 4) Enter the following as theme keywords “Gulf of Maine Ocean Data Partnership,” and “GoMODP”.
- 5) Enter “Gulf of Maine” as a place/location keyword.
- 6) GetData links should bring the user to where the data can actually be obtained – not to the partner’s main page.
- 7) If metadata have already been created and/or registered:
 - Ensure that the keywords suggested in items 4 and 5 above are added and that GetData links go to where the data can be obtained. Notify Melanie Meaux of Global Change Master Directory (GCMD) (Melanie.F.Meaux@nasa.gov) to include your metadata records in the GCMD portal for the Partnership. For partners with FGDC-compliant metadata records in XML format, these can be converted to DIF by sending as an email attachment to Melanie Meaux (Melanie.F.Meaux@nasa.gov).
 - If adding required keywords would be too burdensome due to the number of metadata records, Ms. Meaux may be able to assist.

If creating/registering metadata for the first time:

- 1) Partners should consider the GoMODP Discovery Portal (<http://gcmd.nasa.gov/portals/gomodp/>) on the Global Change Master Directory (GCMD) using the Directory Interchange Format (DIF). This format may be easily converted to FGDC format using online tools. This portal is recommended due to ease in creating and registering metadata. See <http://gcmd.nasa.gov/User/authoring.html>
- 2) Canadian partners are encouraged to register either with the GCMD or with the GeoConnections Discovery Portal (GCP) (<http://www.geoconnections.org/CGDI.cfm/fuseaction/home.welcome/gcs.cfm>). Data registered with either site will be mirrored on the other.
- 3) Registering with the GeoSpatial One-Stop portal (<http://www.geodata.gov>) is also an option. Since GCMD does not actively harvest from GeoSpatial One Stop, if choosing this route please contact Melanie Meaux (Melanie.F.Meaux@nasa.gov) so that she can harvest the metadata for incorporation in the GCMD GoMODP portal.

General recommendations for each metadata registration tool/portal:

- 1) If using the Global Change Master Directory Doc Builder tool, select “Gulf of Maine” for a Location Name and “Gulf of Maine Ocean Data Partnership” as a Project.
- 2) If using GeoConnections:
 - Do not work on metadata outside of the portal (i.e. saving metadata files locally as txt, xml, or html, editing outside of the GeoConnections interface, and uploading again). Problems were experienced in 2005, although this may have been remedied by the portal managers.
 - To save time in registering multiple similar metadata records, copy your initial metadata record by using the “Save As” button and then entering and saving a new Product Collection Name. A new entry with a new ID will be created under your account. This should save you from having to retype everything from scratch if an entry that could be used as a template already exists under your account.
 - If the template entry exists under another account you can be granted permission to make a copy of it or the Discovery Portal can provide you with a copy.
- 3) Registering Data with Geospatial OneStop
 - There are several ways to publish metadata to Geospatial OneStop. Partners should follow the guidance in the publication "Creating and Publishing Metadata in Support of Geospatial OneStop and the NSDI" (<http://www.geodata.gov/gos/metadata/CreatePublishMetadata.pdf>). Additional information can be found at the Geospatial One-Stop web page at <http://www.geodata.gov>. For partners who have existing metadata in the GCMD and would like it to be also published in GOS, please contact Melanie Meaux (Melanie.F.Meaux@nasa.gov) or Lola Olsen (Lola.M.Olsen@nasa.gov)
 - The following is a brief description of the steps required for partners to publish metadata on Geospatial OneStop. Metadata must be created using the FGDC's Content Standard for Digital Geospatial Metadata (CSDGM) XML format (see <http://www.fgdc.gov/metadata>). Tools are available at Geospatial OneStop to facilitate this. Metadata must include a document unique ID, a metadata update date, and inclusion of appropriate ISO 19115 keywords. Partners will need to either input metadata directly into Geospatial OneStop or choose a method of metadata harvesting such as publication in a Z39.50 metadata clearinghouse, ArcIMS metadata service, web accessible folder (WAF), or Open Archive

Initiative (OAI) metadata service. A partner must register their metadata at Geospatial One-Stop for it to be harvested.

Other Metadata Creation and Editing Tools

Several other online tools are available for automating creation and editing of metadata:

- Metadata Tools for Geospatial Data: <http://sco.wisc.edu/wisclinc/metatool/>
- Metadata creation/editing software - USGS: <http://geology.usgs.gov/tools/metadata/>
- Consolidated metadata tools available on the Marine Metadata Interoperability website: <http://marinemetadata.org/tools/refs/>

Data Transport and Access – Recommended Technologies

The following is a table of recommended distributed data systems and technologies.

Data System/ Technology	Web Site	Generally Recommended for	Brief Description
Ocean Biogeographic Information System (OBIS)	www.iobis.org/	Ocean biogeographic data holders.	Partner keeps the dataset locally and sets up a server that can respond to OBIS queries such as Oracle, MySQL and JGOFS/GLOBEC. This requires "mapping" the dataset to the OBIS Schema and installing a free software package called DiGIR to communicate with the OBIS portal.
Exchange Network	www.exchangenetwork.net	State agencies	A secure internet and standards based approach for exchanging environmental data, being developed by the EPA and state environmental agencies. Exchange Network partners establish servers called network nodes that are securely connected to the internet. A node serves as the exchange point for all data requests and submissions. XML (Extensible Markup Language) is employed to transmit data between nodes.
Open Geospatial Consortium (OGC) services	www.opengeospatial.org/	Geospatial data holders	This is a non-profit, international, voluntary consensus standards organization that is leading the development of standards for geospatial based services. Partners who will be contributing geospatial data should consider the OGC specifications such as the Web Mapping Service (WMS) for allowing online access to maps and the Web Feature Service (WFS) for allowing online access to geospatial data.
Generic Web Services	www.w3c.org	Partners not currently using another protocol. Currently may be the easiest to use for partners with limited IT resources.	The World Wide Web Consortium is developing specifications, guidelines and tools for a variety of technologies that fall under the heading of web services. Web service protocols tend to be very general, enable an incremental approach to interoperability that facilitates collaboration, and have widespread use in other disciplines. Other technologies are either already based on XML-based web services (e.g., the Exchange Network, OBIS & OGC) or developing web service interfaces (e.g., OPeNDAP).
Open-source Project for a Network Data Access Protocol (OpenDAP)	http://opendap.org/	Gridded data holders or those with complex data collections and large collections of individual files that comprise a single logical data set.	OPeNDAP allows you to access remote data over the internet using familiar data analysis and visualization packages like Matlab, Ferret, IDL, and other clients. In this way, data are brought directly to your working environment according to your specifications/query. On the server side, data can be in netCDF, ASCII, relational databases, and a variety of other formats.
U.S. Joint Global Ocean Flux Study/U.S. GLOBAL Ocean Ecosystems Dynamics (JGOFS/GLOBEC)	http://globec.whoi.edu/globec-dir/data_org_ret.html	Oceanographic data holders	Used to provide on-line access to physical, chemical and biological data, it has a data object paradigm to allow numeric, image, video and text-based data to be transmitted over the internet. The system uses standard internet browsers and http to serve data and information to all types of computing platforms.

Interoperability

- 1) The groundwork for data interoperability is presented by the standards and practices described in the data authority, metadata and access/transport sections above.
- 2) Partners should participate in the pilot projects where appropriate.
- 3) More guidance for this area will be developed as experience is gained.