

# **Gulf of Maine Ocean Data Partnership**

## **2005 Technical Guidance for Data Sharing**

**December 2005**

### **Technical Committee Members:**

Deb Soule, Technical Committee Chair – New Hampshire Department of Environmental Services

Anne Ball – NOAA – Coastal Services Center

Seth Barker – Gulf of Maine Council on the Marine Environment

Philip Bogden – Gulf of Maine Ocean Observing System

Bob Branton – Bedford Institute of Oceanography

Bob Groman – Woods Hole Oceanographic Institution

Stephen Hale – EPA, Atlantic Ecology Division

Jim Manning – NOAA, Northeast Fisheries Science Center

Just Moller – Stellwagen Bank National Marine Sanctuary

Sylvia Most - Gulf of Maine Ocean Observing System

David Mountain - NOAA – Northeast Fisheries Science Center

Joan Palmer - NOAA – Northeast Fisheries Science Center

Chris Polloni – USGS, Woods Hole Science Center

Annette Schloss – Coastal Ocean Observation and Analysis - UNH

Tom Shyka - Gulf of Maine Ocean Observing System

Lou Van Guelpen - Huntsman Marine Science Centre/Atlantic Reference Centre

## Table of Contents

Table of Contents	2
Executive Summary	3
Background	5
Purpose of Document	5
Obtaining Help – Partnership Assistance	6
Partnership Survey	6
Technical Guidance	7
Desired Datasets – Type and Geographic Scope of Data Collections	7
Data Assurance and Authority	7
Recommended Vocabulary (Standardized Names)	8
Species Names	8
Parameter Names	9
Gazetteer Names	9
Data Discovery – Constructing and Registering Metadata	10
Metadata Creation and Editing Tools	12
Data Transport and Access	12
Where we are today	13
Interoperability	14
Next Steps	14
References	15
Appendix I: Acronym List	16
Appendix II: How accessible are the Partnership’s data? – GoMODP Survey Results	17

## Executive Summary

The Gulf of Maine Ocean Data Partnership (Partnership) is an international effort to promote and coordinate the sharing, linking, electronic dissemination, and use of data from the Gulf of Maine region. Members of the Partnership include governmental agencies, intergovernmental organizations, and academic, research, and other nonprofit entities. Each member organization is engaged in the collection of physical, biological, chemical, or geological data on the Gulf of Maine. The Partnership is primarily interested in multi-year or relatively stable datasets with a surface water/ocean component within the Gulf of Maine watershed area.

This document was created by the Partnership's Technical Committee to provide guidance and specific recommendations in the five key areas needed for successful data sharing and interoperability. Guidelines for *ensuring data assurance and authority, recommended vocabulary, constructing and registering metadata, data transport and access, and interoperability* are presented as of December 2005. These guidelines are expected to evolve as technology and Partnership concerns change. The major recommendations are listed below:

### Data assurance and authority:

- The Partnership prefers datasets where ownership is established and the partner is fully authorized to share the data.
- Partners should properly preserve/archive their data sets.

### Recommended vocabulary (standardized names):

- Where possible, partners should utilize industry-standard naming conventions within their dataset. If a non-standard list is used, a crosswalk between the non-standard list and a standard list may need to be developed.
- For partners not yet committed to a source of standard species names, the Integrated Taxonomic Information System (It is) is recommended.
- The Marine Metadata Interoperability website provides sources for standard lists of parameter names.
- For place names, partners are encouraged to use the Geographical Names of Canada or the USGS Geographic Names Information System (GNIS).
- Crosswalks are already under development in some areas to enable datasets using different naming conventions to be combined. The Partnership will look to use/expand on these efforts.

### Constructing and registering metadata:

- Partners are encouraged to construct metadata in FGDC format or use the DIF format (which can be converted to FGDC format) through the NASA's Global Change Master Directory (GCMD) web site.
- Each partner should register their own metadata. The Partnership will provide assistance to partners in constructing and registering metadata.
- Portals such as GeoConnections (Canada) and GCMD (USA) are preferred, as any metadata registered on either of these sites will be harvested by the other site. Partners already registered with the GeoSpatial OneStop web site are requested to contact the GCMD to harvest their metadata records.
- Partners are encouraged to enter the following as theme keywords "Gulf of Maine Ocean Data Partnership," and "GoMODP". "Gulf of Maine" should be used as a place theme word where applicable.

### Data transport and access:

- Partners are currently utilizing a variety of technologies for data sharing with customers and should not stop using any technologies that are currently in place.

- For partners beginning data sharing efforts, a list of recommended distributed data systems and technologies are provided, including OBIS, the EPA Exchange Network, OGC Services, OpenDAP and JGOFS/GLOBEC.

Data interoperability:

- The Partnership will develop pilot projects to demonstrate standards for interoperability and integrated data products.

## Background

The Gulf of Maine Ocean Data Partnership (Partnership) is an international effort to promote and coordinate the sharing, linking, electronic dissemination, and use of data from the Gulf of Maine region. Members of the Partnership include governmental agencies, intergovernmental organizations, and academic, research, and other nonprofit entities. Each member organization is engaged in the collection of physical, biological, chemical, or geological data on the Gulf of Maine.

The Partnership is a coordinated effort to enable users throughout the Gulf of Maine region to discover and put to use vast and growing quantities of data in databases housed at agencies throughout the northeastern United States and Canada. Through coordinated access to respective databases, partners wish to advance a truly integrated ocean observing system in the Gulf of Maine.

To accomplish this, the Partnership chose to develop a set of recommendations to ensure that all partners could serve their data via the Internet. Also, the first year's work plan suggests the challenge of interoperability would be tackled by stressing standardization within disciplines and use of existing protocols for data sharing, rather than creating new standards.

At the December 7, 2004 meeting of the Partnership, a formal Technical Committee was created to provide its members with guidance and specific recommendations in five key areas needed for successful data sharing and interoperability. These areas include

- Recommend criteria to establish *data authority*
- Recommend *standard names* for species, parameters and localities
- Provide guidelines for *constructing metadata*
- Recommend *data transport and access* methods
- Recommend "best practices" to facilitate *data interoperability*

Throughout 2005, the Technical Committee has met to address these five key areas. In addition to drawing upon its members' individual experiences and expertise, the technical committee took advantage of a metadata pilot study (project report is available at [www.gomodp.org](http://www.gomodp.org)), an online survey to partners in spring 2005 and other documentation including Hankin et al., 2005 and the Report of the Study Group on Management of Integrated Data (SGMID), referenced at the end of this document.

This Partnership is part of a larger effort -the regional associations of the Integrated Ocean Observing System (IOOS), the national IOOS, and the international Global Ocean Observing System (GOOS). All the regional associations and the IOOS Data Management and Communications group are struggling with the same data issues. Solutions and common standards are emerging and will change over the next few years. This document is interim guidance and is anticipated to be updated yearly or as significant new projects develop.

## Purpose of Document

The purpose of this document is to provide guidance and specific recommendations in the five key areas needed for successful data sharing and interoperability. Each area is described in greater detail below.

*Data authority:* Data authority references who owns data and whether the data contributor has authority to share them with others. This is important to ensure that rights of the original data collector(s)/owner(s) are protected, and to minimize risks of serving outdated information or serving the same data more than once.

*Standard names for species, parameters and localities:* Where standards have been developed for identification of species, parameters or place names (gazetteers), the use of these standards facilitate communication among partners and interoperability of datasets.

*Constructing and registering metadata:* Metadata, or information about the data, is critical to the data discovery phase so that users querying available web portals can correctly identify data sets of interest. In addition, metadata enable users to make full use of data on the web, providing descriptions of instruments and techniques used in making measurements, and processing steps employed to transform raw measurements.

*Data transport and access:* Data access methods provide means for exchanging data and information among disparate systems. These methods are as simple as exchanging floppy disks or downloading a file via ftp. But to enable computer-to-computer interoperability, these methods need to be based on digital exchange protocols such as Open-source Project for a Network Data Access Protocol (OPeNDAP) and Open Geospatial Consortium (OGC) Web Services, or via direct access to data management systems such as Oracle.

*Data interoperability:* Data interoperability requires a certain level of data consistency in terms of naming conventions, standard vocabularies and content specifications. This is also important during the data discovery or querying phase when people are trying to locate data of interest. As mentioned before, recommending standard species, parameter and locality names are an important part of this effort.

Lastly, best practices recommendations provide suggestions, hints, and approaches to data storage, data access and data retrieval that can facilitate data interoperability.

The sections below elaborate on the issues, and provide suggestions and recommendations to the partners to best reach the goal of serving and sharing their data over the Internet.

## **Obtaining Help – Partnership Assistance**

The Partnership exists to not only share data but also to help its individual partners achieve the necessary capabilities to share data effectively and efficiently without undue burden. To that end, a partner having difficulty meeting criteria and recommendations in this document is encouraged to contact the Partnership's host, the Gulf of Maine Ocean Observing System (GoMOOS), at [www.gomoos.org](http://www.gomoos.org). To provide and encourage assistance between partners, a document will be compiled in early 2006 which will list each partner's data expertise. This document will be posted along with all the other technical guidance documents at the Partnership's website: [www.gomodp.org](http://www.gomodp.org).

## **Partnership Survey**

To aid in developing technical guidance for the Partnership, partners were asked in spring/summer 2005 to update and expand on information provided in a previous survey using an online tool developed for our wiki site. The revised survey asked more detailed questions in the area of standards, data sharing methods and institutional constraints. In addition to helping generate guidance, the questions were intended to help partners generate a minimum FGDC-compliant metadata record for their dataset. Responses to the survey

were mixed, and compilation of the results was limited by the fact that some of the larger partners with full metadata records already published provided links to their records rather than filling out the survey. These responses did not get incorporated in the survey results. This represents one limitation of this analysis, as we have incomplete data in this report and “no response provided” type entries in a number of key areas. Some results are summarized in the main body of this document while the rest are in Appendix II.

## **Technical Guidance**

### **Desired Datasets – Type and Geographic Scope of Data Collections**

Initially, the Partnership is primarily interested in multi-year or relatively stable datasets (such as one-time eel grass mapping) with a surface water/ocean (not groundwater) component within the Gulf of Maine watershed area. An outer ocean boundary has not been explicitly stated. However, important datasets for areas outside this watershed, such as Nova Scotia and south of Cape Cod, will not be excluded. Using the watershed as a defining basis initially focuses data compilation/sharing efforts. 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**

The Partnership assumes that each partner is ultimately responsible for quality assurance and quality control of data they share. The Partnership will recognize and respect ownership rights for data provided by each partner. 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. Determining who is the owner or has the authority to serve a dataset can sometimes be tricky. For example, a dataset may be collected in collaboration with other organizations, or one group may acquire a dataset from another and make enhancements to it. 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:
  - i. Are you the primary creator/collector of the dataset or is there shared ownership?
  - ii. If you are not the primary creator/collector, do you have a formal distribution agreement with the data provider?
  - iii. Who else has the data or is serving them?
  - iv. Who will correct errors in the event any are discovered?
  - v. 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 network and data backup offsite to avoid loss in the case of a local event.
- 6) If a partner can not 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) To the extent that 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)

Sharing data, developing interoperability, and retrieving accurate results via a query can become difficult when different terminology is used across datasets, particularly with regard to names for species, environmental parameters, or localities. Organizations are not precluded from developing their own terminology, however, the Partnership recommends adoption of standardized names from sources recommended below. If developing their own terminology, it is recommended that partners consult these sources of standard names.

To investigate the variety of naming conventions used by partners for species, parameter, and place names, the revised survey was enhanced to include questions about these items. It was hoped that these survey results would help in developing guidance for recommended vocabularies. The survey results (as of October 1, 2005) clearly demonstrated how difficult this issue will be to resolve (since there was no majority of use for any particular standardized names) and how much more work and discussion are needed in this important area. The following section describes survey results for each of the three areas (species, parameter, and gazetteer) and suggests next steps.

## Species Names

In the survey, partners were asked to describe what standardized source of species names was used (if any) for each dataset they reported. Several common sources for species names, such as: American Fisheries Society's Common and Scientific Names of Fishes from the United States and Canada, Integrated Taxonomic Information System, Fishbase, Species 2000 etc. were provided as check box options. Partners were also given the option of declaring if another or no list was used. Here are the survey responses for the 79 datasets described:

Standardized Species List Used	Number of Datasets
Integrated Taxonomic Information System (ITIS)	2
Fishbase	1
Other Species List	42
None in use	7
No response provided	28

One survey response indicated that 2 different lists were used; therefore there is a total count of 80 datasets in the table and not 79. The lack of response for 28 datasets may indicate that species were not part of the data (i.e., physical/chemical data could have been the focus of the dataset). For the 42 in which "Other Species List" was selected, no description was provided of the naming convention used.

One resolution to differing species names is for data contributors to develop a vocabulary "crosswalk" or matrix that matches the taxonomic terminology among the different partners. However, this is a very difficult undertaking since species names have changed over time, both to correct previous identifications and because of developments in taxonomy. Several efforts are underway that may help the Partnership with this problem. For example, the Universal and Biological Indexer and Organizer project (UBIO) provides a Taxonomic Name Server (see <http://www.ubio.org/SOAPbrowser/>) that may provide the crosswalk functionality we require in the form of an automated service, at least on a limited basis.

For those partners not yet committed to a source of standard names, ITIS is recommended. ITIS provides a look-up service for common and scientific names, as well as names of higher taxonomic levels, that functions at the user (consumer) level, in contrast to the UBIO service that will ultimately provide a machine-to-machine crosswalk.

### Parameter Names

For the purposes of this effort, the term parameters refers to those variables measured in the field or analyzed in a lab not already described under the species list. Partners were asked to describe the standardized source of parameter names they used for each dataset by selecting from a check list of "British Oceanographic Data Centre" or "Other". If "Other" was selected, partners were asked to declare what source was used.

There were 50 "Other" responses and only 2 descriptions were provided: Climate and Forecast metadata convention (CF) and the US Environmental Protection Agency's (EPA) STORET database in addition to the Substance Registry System (also maintained by EPA). The lack of responses (29 in total) could indicate that only non-parameter data were collected in the dataset; therefore no standard naming convention for parameters was referenced or a proprietary list of names was used.

Once again there is no common standardized source used by a majority of partners and the Partnership is faced with a situation similar to that for species names. Unlike species names, a vocabulary crosswalk may be easier to develop or obtain by the Partnership. However, this work may be outside the resources of the Partnership.

The Partnership may be able to use/expand on crosswalks under development/listed by the Marine Metadata Interoperability web site (<http://marinemetadata.org/vocabularies/refs/mapping/>). While the Partnership pursues a solution to this situation, partners are encouraged to use one of the sources mentioned on the Marine Metadata Interoperability web site (<http://marinemetadata.org/vocabularies/refs/markup/>) rather than developing new parameter lists for their datasets.

### Gazetteer Names

While it is desired that all datasets contain spatial coordinate data and therefore allow for spatial queries, it is also useful to have the data referenced to common geographic names for non-coordinate queries. In addition, representative coordinates (such as the geographic center of a bay) can be obtained for these standardized place names. Partners were asked to select from a short list of known sources of place names or specify if some "Other" list or no list was used/consulted for each dataset. Here are the responses:

Standardized Gazetteer List Used	Number of Datasets
USGS Standard Names	6

Canadian Geographic Names	2
No list used	7
Other list	6
No response	66

One response indicated that both the USGS Geographic Names Information System (<http://nhd.usgs.gov/gnis.html>) and Geographical Names of Canada ([http://geonames.nrcan.gc.ca/index\\_e.php](http://geonames.nrcan.gc.ca/index_e.php)) were used for a dataset. Another respondent stated that for 5 datasets both the USGS and another list were used when the USGS did not provide a place name.

Once again, there is no majority use of any particular gazetteer. The Partnership may need to select a standard and develop a crosswalk of partners' place names or coordinates (if provided) to the selected standard, or search/wait for the development of a crosswalk by an outside party. In the beginning, the Partnership may decide to accept only coordinates for spatial queries to avoid additional crosswalks at this time. Or, partners may choose to use a gazetteer such as those mentioned above to produce coordinates in a database where none were given originally. In this case, documenting the choice through metadata will be essential. If no standard is presently in use, partners are encouraged to use the Geographical Names of Canada or the USGS Geographic Names Information System (GNIS) at <http://nhd.usgs.gov/gnis.html>.

### **Data Discovery – Constructing and Registering Metadata**

Many organizations have developed great datasets which would be helpful to other organizations - if they were only aware of them! While some partners have already made their datasets discoverable by registering their metadata with one of several public discovery portals, many partners have not and would likely need technical assistance to do so.

As an easy first step in making the partners' datasets known or discoverable to the entire Partnership, a dataset inventory was developed by compiling responses to a detailed survey. The results were then posted on the Partnership's web site at [www.gomodp.org](http://www.gomodp.org), under the topic "Partnership Data Survey."

Another step to further data discovery was taken when a pilot project was conducted to assist 5 partners in constructing and registering their metadata with the GeoConnections Discovery Portal. Lou Van Guelpen of the Huntsman Marine Science Centre assisted the NOAA-Northeast Fisheries Science Center, Maine Department of Marine Resources, Stellwagen Bank National Marine Sanctuary, Tufts University SEANET, and the Wells National Estuary Research Reserve in this endeavor. At the conclusion of the pilot project, a report (available at [www.gomodp.org](http://www.gomodp.org)) was submitted detailing steps taken, obstacles encountered, and recommendations for extending registration of metadata to other partners. The pilot project demonstrated that personal guidance to non-registered partners is likely needed for the registration effort to proceed efficiently. In addition to a metadata training session for the partners, partners will likely require guidance from a designated resource person or organization during metadata registration. Therefore, a similar service to the pilot project will be offered in the future for partners.

For Partnership purposes, the following are general guidelines for developing 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) If metadata have already been registered:
  - a. Ensure that the keywords suggested in items 4 and 5 above are added.
  - b. Notify Melanie Meaux of Global Change Master Directory (GCMD) ([mmeaux@nasa.gov](mailto:mmeaux@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 ([mmeaux@nasa.gov](mailto:mmeaux@nasa.gov)).
  - c. If adding required keywords would be too burdensome due to the number of metadata records, Ms. Meaux may be able to assist using GCMD functionality

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 ([mmeaux@nasa.gov](mailto:mmeaux@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 doing this in the summer of 2005, although this may be 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 could either be granted permission to make a copy of it or the Discovery Portal can provide you with a copy.
- 3) Recommendations for using the GeoSpatial One-Stop are not available at this time and will be developed in the coming year.

## **Metadata Creation and Editing Tools**

Several 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/>
- Global Change Master Directory on-line tool: <http://gcmd.nasa.gov/User/authoring.html>
- Consolidated metadata tools available on the Marine Metadata Interoperability website: <http://marinemetadata.org/tools/refs/>

The newly formed IOOS DMAC Expert Team on Metadata and Data Discovery is an ongoing source of information on metadata tools. The Partnership will look to this group for future guidance.

## **Data Transport and Access**

An overarching goal of the Partnership is to enable the partners and other stakeholders to dynamically share and integrate distributed data sets. In order to achieve this ‘interoperability’ the Partnership will need to define architecture with standards for data access and transport.

The Partnership includes a variety of organizations that collect, manage and serve a wide range of environmental data and information products. The partners also use a variety of technologies to accomplish data transport and access (see table below). Some of the technologies used by the partners were developed for a specific data type (e.g. OGC web map service) while others are even more generic (e.g. Simple Object Access Protocol (SOAP) service) and could be used for many data types.

It is recommended that the Partnership attempt to minimize the number of technologies used for data access and transport. Defining a common set of technologies will allow the Partnership to develop or leverage existing interfaces to make connections between a relatively small set of data access technologies, thus allowing application-to-application communication and interoperability. A set of existing technologies that are adequate for most data sets and that will increase interoperability will be recommended. This list of technologies will evolve as technologies evolve.

Exchange protocols in use	Number of Datasets
OPeNDAP	10
OGC Web Map Service (WMS)	7
OGC Web Feature Service (WFS)	1
FTP	7
HTTP/HTTPS	3
EPA Exchange Network	1
Other/specify (most common alternative was CDROM)	15
DiGIR	17
No response provided	39

Progress can and should be incremental. The Partnership should look to define test bed activities with limited scope to demonstrate progress with an end-to-end solution that can be further developed in an incremental fashion. Metadata and data schema issues will be raised and can likewise be addressed incrementally.

The Partnership should:

*Continue to serve customers:* Partners should not stop using any technologies that are currently in place. The Technical Committee may recommend migrating to a new technology, will help assess the feasibility, advantages and disadvantages of doing so, and will try to provide technical support as necessary.

*Develop a web services approach:* Web services and service-oriented architectures seem to be emerging as the foundation for most approaches. Specific guidance and pilot projects will be developed for the Partnership to move toward web services architecture and existing standards will be leveraged wherever appropriate.

## Where we are today

The following is a list of distributed data systems and technologies that are recommended. Partner organizations are using one or more of the systems and technologies listed below. The Partnership will develop a resource list of contacts within the Partnership that have expertise with the recommended systems and technologies. The GoMODP host, GoMOOS, will provide technical assistance to the partners with some of the technologies.

*Ocean Biogeographic Information System (OBIS; [www.iobis.org/](http://www.iobis.org/)):* Ocean biogeographic data holders should consider becoming an Ocean Biogeographic Information System (OBIS) distributed data contributor. The partner will keep the dataset locally, and sets up a server that can respond to OBIS queries. This requires "mapping" the dataset to the OBIS Schema and installing a free software package called DiGIR to communicate with the OBIS portal. Today OBIS supplies thousands of records on the Gulf of Maine.

*Exchange Network ([www.exchangenetwork.net](http://www.exchangenetwork.net)):* The Exchange Network is a secure Internet and standards based approach for exchanging environmental data, which is being developed by the EPA and State environmental agencies. Partners on the Exchange Network establish and maintain servers called network nodes that are securely connected to the Internet. A node is a partner's single point of presence on the Exchange Network and serves as the exchange point for all data requests and submissions. The exchange network employs XML (Extensible Markup Language) to transmit data between nodes.

*Open Geospatial Consortium (OGC) services ([www.opengeospatial.org/](http://www.opengeospatial.org/)):* The Open Geospatial Consortium, Inc. (OGC) is a non-profit, international, voluntary consensus standards organization that is

leading the development of standards for geospatial and location 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:* The World Wide Web Consortium ([www.w3c.org](http://www.w3c.org)) is developing specifications, guidelines and tools for a variety of technologies that fall under the heading of Web services. These technologies are receiving increasingly wide acceptance in a variety of fields. GoMOOS has begun using the Simple Object Access Protocol (SOAP), a particularly popular W3C XML protocol, for a variety of purposes. Web service protocols, such as SOAP, tend to be very general, and they have been implemented with software tools on every platform in practical use. While they are not a panacea, they enable an incremental approach to interoperability that facilitates collaboration. Technically, because of their widespread use in other disciplines, they may also represent the easiest tool set for partners to adopt. Moreover, 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, see below).

*Open-source Project for a Network Data Access Protocol (OpenDAP; <http://opendap.org/>):* Data providers who will contribute gridded data should consider providing access through the OPeNDAP access protocol. Other data types appropriate for OPeNDAP include complex data collections in a relational database and large collections of individual files that comprise a single logical data set.

*JGOFS/GLOBEC:* The JGOFS/GLOBEC data management software (overview available at <http://globec.whoi.edu/globec-dir/doc/datasys/jgsysfull.shtml>) is used by the U.S. JGOFS and U.S. GLOBEC programs as well as several other oceanographic projects to provide on-line access to their physical, chemical and biological data. The JGOFS/GLOBEC system uses a data object paradigm to allow numeric, image, video and text-based data to be transmitted over the Internet. The system uses standard browsers (e.g., Internet Explorer, Netscape, Mozilla, etc.) and the hypertext transmission protocol to serve data and information to all types of computing platforms. Planned enhancements will simplify computer-to-computer interoperability by using FGDC/XML to exchange metadata information.

## **Interoperability**

Another of the overarching goals of the Partnership is to enable the use of various data sets by the partners and stakeholders in the region to solve problems and further our understanding of the natural systems in the Gulf of Maine. This will require the integration of distinct and distributed data sets. The groundwork for data interoperability are presented by the standards and practices described in the data authority, metadata and access/transport sections above. Full data interoperability is a long term aim but short term success is achievable by using existing standards.

The Partnership will develop interoperability pilot projects to demonstrate how standards can enable interoperability and integrated data products. The partners are encouraged to participate in the pilot projects where appropriate.

## **Next Steps**

Interoperability of data between partners is the ultimate goal of the Partnership and activities for 2006 will be geared toward achieving that goal. Activities will include (but not be limited to):

- Achieving a greater percentage (ultimately 100%) discoverability of partners' data through constructing and registering metadata.
- Creating a table describing partner's data expertise in areas such as constructing metadata, registering metadata, exchange protocols, etc.
- Monitoring existing efforts to develop crosswalks between disparate vocabularies and develop crosswalks only as a last resort.
- Piloting projects to demonstrate interoperability between disciplines.

Other activities are likely to be identified through the 2006 work plan process.

## References

Hankin, S. and the DMAC Steering Committee, 2005, Data Management and Communications Plan for Research and Operational Integrated Ocean Observing Systems: I. Interoperable Data Discovery, Access, and Archive, Ocean.US, Arlington, VA 304pp. [http://dmac.ocean.us/dacsc/imp\\_plan.jsp](http://dmac.ocean.us/dacsc/imp_plan.jsp)

International Council for the Exploration of the Sea (ICES). Report of the Study Group on Management of Integrated Data April 11-13, 2005. Available at: <http://www.ices.dk/reports/ACE/2005/SGMID05.pdf>

## Appendix I: Acronym List

CGDI	Canadian Geographic Data Interchange
DIF	Directory Interchange Format
DiGIR	Distributed Generic Information Retrieval
EPA	Environmental Protection Agency
FGDC	Federal Geographic Data Committee
FTP	File Transfer Protocol
GCMD	Global Change Master Directory
GCP	GeoConnections Discovery Portal
GLOBEC	GLOBal Ocean ECosystems Dynamics
GNIS	Geographic Names Information System
GoMODP	Gulf of Maine Ocean Data Partnership
GOOS	Global Ocean Observing System
HTTP	Hyper Text Transfer Protocol
IOOS	Integrated Ocean Observing System
ITIS	Integrated Taxonomic Information System
JGOFS	Joint Global Ocean Flux Study
MMI	Marine Metadata Interoperability
NEIEN	National Environmental Information Exchange Network
NOAA	National Oceanic and Atmospheric Administration
OBIS	Ocean Biogeographic Information System
OGC	Open Geospatial Consortium
OPeNDAP	Open-source Project for a Network Data Access Protocol
SEANET	Seabird Environmental Assessment Network
SGMID	Study Group on Management of Integrated Data
SOAP	Simple Object Access Protocol
STORET	EPA's STOrage and RETrieval database
UBIO	Universal and Biological Indexer and Organizer
USGS	United States Geologic Service
W3C	World Wide Web Consortium
WFS	Web Feature Service
WMS	Web Mapping Service
XML	Extensible Markup Language

## Appendix II: How accessible are the Partnership’s data? – GoMODP Survey Results

Through the partnership survey, we attempted to find the current status of metadata registration by partners. Most datasets documented through this survey have FGDC compliant metadata records, yet only a third of them are available online.

Do you have an FGDC compliant metadata record for this dataset?	Number of Datasets
No	11
No, other standard used, please specify	4
Yes (Provide URL if available)	61 (22)
No response provided	4

Results for the question of how metadata are published were inconclusive. Many respondents selected “other-specific” without specifying an alternate method. This implies an error in the wording of the question and a need for further inquiry.

If you publish your metadata, do you serve them using any of the following?	Number of Datasets
Not published	9
In house Z39.50 server	2
External Z39.50 server	0
Other/Specify	34
No response provided	34

Accessibility of the datasets themselves presented different issues. Through the survey, partners were asked about data access constraints, such as firewall issues and institutional constraints. Few partners listed firewall issues as a concern. Nearly two-thirds of the datasets are currently available from online locations. Additionally, there appear to be few institutional constraints to the sharing of data. All of these are positive signals for the ability of the Partnership to pursue data sharing arrangements.

Dataset is available from an online location?	Number of Datasets
No	24
Yes (specify URL)	48 (45)
No response	7

Technical constraints to use of this dataset	Number of Datasets
Firewall issue	5
Non-standard TCP service port	1
Other/Specify	5

Institutional constraints to use of this dataset	Number of Datasets
No constraints	6
Data freely available with a standard disclaimer	20

Data freely available after permission granted from data provider	2
Data freely available, however appropriate recognition must be provided	40
No response provided	11

The survey attempted to survey the variety of options available to partners in databases and file types. There were a wide range of responses, with no predominant type.

Database types (check all that apply)	Number of Datasets
Oracle	11
PostgreSQL	2
MS-Access	7
MySQL	3
MS SQL Server	1
ESRI SDE	3
Informix	6
DB2	1
Other/specify (predominant alternative was MS Excel)	11
No response provided	47

File Formats available (check all that apply)	Number of Datasets
CSV	9
Tab-delimited	6
ASCII	7
ESRI Shapefile	4
ESRI Export (E00)	0
Geo-referenced raster (GeoTIFF, PNG, JPEG)	4
Other/specify *	29
No response provided	36

\* Multiple "other" alternatives were provided, such as HDF-EOS, netCDF, Arc-ASCII GRID, BSQ, text file, HTML, KGSMapper, C-square Mapper, Easy Netviewer, ERDAS Imagine (.img) format, Arc/Info coverage format, SDTS