Report of the First HyCODE Data Management and Distribution Workshop

(Version: June 11, 1999)

Rutgers University

April 12-13, 1999


1.Ocean Physics Laboratory, University of California, Santa Barbara, 6487 Calle Real, Suite A, Goleta, CA 93117, email: tommy@ices.ucsb.edu
2.Office of Naval Research
3.Naval Research Laboratory
4.Florida Environmental Research Institute
5.University of South Florida
6.Rutgers University
7.PSI
8.Johns Hopkins University/Applied Physics Laboratory

Contents

Workshop Report

1. Background
2. Introduction to Workshop
3. Discussion Items
   a. Description of Field Data Sets
   b. Satellite and Aircraft Data and Algorithms
   c. Modeling Data Policy
   d. Special Navy Requirements and Transitions
   e. The Rutgers LEO-15 RODAN Data System
   f. The WOOD System
   g. Data Formats
   h. Data Policy Issues
   i. Summary of General Issues
   j. Timeline of Actions
   k. Assignment of Tasks

Appendices

I. Participants
II. Workshop Agenda
III. David Porter Message
IV. Tables
1. **Table 1.** Data availability, observational platforms, and instrumentation for LEO-15
2. **Table 2.** Data availability, observational platforms, and instrumentation for west Florida shelf (WSF)
3. **Table 3.** Data availability, observational platforms, and instrumentation for CoBOP
4. **Table 4.** Required Combinations of Data Type & Header Attributes

V. Figures

1. **Background**

The Office of Naval Research (ONR) Hyperspectral Coupled Ocean Dynamics Experiments (HyCODE) program is a five-year ONR initiative, which began in January 1999. HyCODE is designed to utilize hyperspectral imagery to improve understanding of the diverse processes controlling inherent optical properties (IOPs) in the coastal ocean. The program will also develop operational ocean color algorithms in both the optically-shallow ocean (i.e., where the signal detected by the satellite sensor is affected by the seafloor) and the optically-deep ocean (i.e., where the signal detected by the satellite sensor is not directly affected by the seafloor). HyCODE field experiments will be focused on three locations: 1) off the coast of New Jersey at the LEO-15 site, 2) on the west Florida shelf (WSF) off St. Petersburg, FL, and 3) in the Bahamas off Lee Stocking Island (LSI) as part of the ONR CoBOP program. A broad range of advanced optical and physical measurements and models will be used for HyCODE.

HyCODE activities are linked to planned Navy satellite sensors with advanced color sensing capabilities. In conjunction with private industry, the Department of the Navy is developing a satellite-based hyperspectral ocean color imager with a planned launch date in the year 2000. Details concerning the satellite, Navy Earth Map Observer (NEMO) are presented below in Section 3b. The Naval Research Laboratory (NRL) has responsibility for calibration; development of at-launch algorithms; on-board data compression software and related software for NEMO’s ground data system; some data processing; and data archiving. Products of at-launch algorithms will include: water clarity (diffuse attenuation coefficient at 490 nm), concentration of chlorophyll, absorption coefficient of colored dissolved organic matter, a measure of suspended sediments, bathymetry, and bottom characteristics. The Navy’s goal for this program is to demonstrate the scientific and technical capability to characterize the littoral environment with remote sensors.

The initial meeting of HyCODE principal investigators was held in Ellicott City, Maryland January 18-21, 1999. It was agreed that data management and distribution was an important aspect of HyCODE. Further, it was deemed essential that planning for data management and distribution begin very early in the program for a variety of reasons. In particular, data are already being collected as part of other ongoing field programs at the three primary HyCODE field study sites: LEO-15 off the New Jersey coast (near Tuckerton, NJ), the west Florida shelf (WSF; off St. Petersburg, FL), and near Lee Stocking Island (CoBOP program; in the Bahamas). Data management for the three sites is being done independently and is at different stages of development. The consensus was that HyCODE data management could build on these ongoing activities and that each could benefit from the experience of the others. Another important program, World Ocean Optical Data (WOOD), is developing an optical database, which can be used for the HyCODE data sets. The focus of HyCODE is on field observations; however, it is critical that these data be synthesized with the NEMO satellite observations for groundtruthing, algorithm development, and scientific studies. For this reason, the HyCODE data planning activity is working closely with those who will be responsible for the satellite and aircraft remote sensing data sets.

2. **Introduction to Workshop**
The first HyCODE Data Management and Dissemination Workshop was held at Rutgers University April 12-13, 1999. This meeting called together several of the key personnel, who will be responsible for various aspects of HyCODE data activities (please see Appendix I for list of participants). An important goal of the meeting was to develop communications among those, who will be responsible for HyCODE’s data and scientific activities. The format of the meeting was informal and a portion of the workshop was spent with the LEO-15 data management group in their computer facility. The Rutgers RODAN data system and several of the pertinent web sites (e.g., WOOD and others) were explored by the group. A demonstration of data telemetry methods was also presented. This was a very productive session in terms of information and idea exchange.

The general objectives of the workshop were to 1) discuss options and plans for the management and dissemination of the HyCODE data sets and 2) develop a plan to facilitate data utilization for the program. The meeting agenda is included as Appendix II.

3. Discussion Items

a. Descriptions of Field Data Sets

The types of data sets, which are being and will be collected at the three HyCODE sites, are summarized below:

i. LEO-15 (Oscar Schofield and Scott Glenn)

Satellite ocean color data supply a synoptic perspective on radiatively active constituents on space and time scales not possible using shipboard sampling. Remotely sensed data primarily provides data on the surface layers of the ocean. Consequently, remote sensing approaches must assume or rely upon ancillary data to estimate vertical changes in the optical properties. In the optically deep ocean (ODO), there is significant vertical variability in the concentrations of phytoplankton, colored dissolved organic matter (CDOM), suspended sediments, and detrital material which can change on the time scale of hours. This is especially true for most nearshore coastal ocean environments. This vertical heterogeneity in the IOPs thus limits the operational and predictive utility of algorithms based solely on satellite ocean color data.

The LEO-15 HyCODE working group, consisting of numerous academic/federal/corporate institutions, is focused on developing, validating, and forecasting algorithms for hyperspectral inherent optical properties in the coastal ocean. Specifically, efforts are focused on developing and validating an integrated adaptive sampling and modeling system for nowcasting and forecasting of the 3-dimensional evolution of inherent optical properties (IOPs) in coastal waters off New Jersey. Efforts will focus on defining the dynamic behavior of nearshore coastal optical properties using the LEO-15 observation network. Data will be used to define the relationships between the nearshore optical properties and the major radiatively active constituents during summer upwelling events. The LEO-15 network is an existing multi-platform network consisting of satellites, radar, autonomous nodes including moorings, research vessels, and autonomous underwater vehicles (AUVs). Table 1 summarizes the observational sampling program and Figure 1 illustrates the sampling region of LEO-15. A timeline of field activities is shown in Figure 2.

ii. West Florida Shelf (Jeff Donovan and Bob Weisberg)

The USF moored array for the west Florida continental shelf derives funding from several different sources. Designed to study inner-shelf processes from a multidisciplinary perspective, the array combines resources from the State of Florida Coastal Ocean Monitoring and Prediction System (COMPS), the NOAA/COP Ecology of Harmful Algal Blooms (ECOHAB) Program, and the Office of Naval Research.
Figure 3 shows the distribution of deployed or planned moorings on the west Florida continental shelf and Figure 4 presents a timeline of WFS activities. The array consists of surface (squares/triangle) and bottom-mounted moorings (x’s) with positions and instruments in correspondence with Table 2. The nomenclature bears on the initiating program; whence, CM, EC, NA denote COMPS, ECOHAB, and ONR respectively. The SE designation, also under ONR auspice, specifies a set of bottom boundary layer arrays for sediment resuspension studies. The surface moorings generally have meteorological instrumentation, a downward looking ADCP for velocity profiles, and several Sea-Bird MicroCATs for temperature and salinity (plus additional temperature sensors). The bottom-mounted moorings have an upward looking ADCP for velocity profiles and a Sea-Bird SBE-26 system for temperature, salinity, and waves. The center of the array is mid-way between Tampa Bay and Charlotte Harbor; two estuarine regions on Florida’s west coast that provide sources of land drainage for the shelf. The array, while spanning the shelf, has most of its resources landward of the 50m isobath, with major concentration about the 25m isobath. The rationale is to resolve the inner-shelf and its responses to the regularly occurring synoptic weather systems. Our measurement goal is to document the evolution of these coastal ocean responses to synoptic weather forcing and to determine the seasonal modulation of these responses owing not only to the seasonal modulation of the wind stress forcing, but also to the seasonality of the background circulation, the density field, and the buoyancy flux inputs by air-sea interactions and land runoff. These same synoptic and seasonal scale changes in the ocean’s dynamical responses are anticipated to provide control (through coupled biological, chemical and geological processes) on spectral inherent optical properties (IOPs). Along with the moored array, the ECOHAB regional field study includes monthly cruises during which we sample hydrographic (temperature, salinity, and nutrients) properties and associated red-tide related biology. Thus, the combined COMPS, ECOHAB, and ONR programs provide an observational background context for HyCODE.

iii. CoBOP/Lee Stocking Island (Weilin Hou)

Coastal Benthic Optical Properties (CoBOP) is a 5-year initiative, started in FY-1997, to investigate optical processes associated with the shallow ocean floor. The science objectives of this initiative are:

* To define and develop the means to measure the inherent optical properties associated with coastal benthic environments.
* To verify state-of-the art radiative transfer models for optically-shallow water in controlled field tests.
* To investigate the relationships between measured benthic optical properties and associated biological, chemical and physical processes.

The initiative is field oriented and emphasizes the interaction of light with coral reefs, sea grasses, and associated marine sediments. The primary field area is in the Bahamas near Lee Stocking Island (Figure 5). The observational assets and instrumentation are summarized in Table 3. A timeline for CoBOP activities is presented in Figure 6. Collaboration is on-going between CoBOP and other research and development programs (within DOD) concerned with remote sensing and underwater imaging. CoBOP is merging into the HyCODE program.

CoBOP information may be found on the following web site: http://www.psicorp.com/cobop and the CoBOP PI email alias is cobop@mit.edu.

b. Satellite and Aircraft Data and Algorithms

Several algorithms that will be coming from the HyCODE program will be used in generating new bio-optical products. The HyCODE algorithm development team is preparing the architecture and
validation procedures for these satellite algorithms. The algorithms are being developed for use on in situ, and satellite data. Satellite data include NEMO and SeaWIFS. NEMO is scheduled for launch in December 2000 and SeaWIFS data are currently available.

i. HyCODE Access to NEMO Data (Bill Snyder)

The following paragraphs describe the basic NEMO scene, the archived data product available for that scene, and information for accessing and analysis of archived data.

**Basic NEMO Scene:**

The NEMO orbit is Sun-synchronous with a 10:30 am equatorial crossing time, ascending node. The satellite will pass over any given site once every seven days, although on average a scene can be re-observed every 2-3 days using cross-track pointing. A basic Navy scene obtained with the COIS VNIR and SWIR sensors will be 30 km wide by 200 km long with 30 m pixel resolution at nadir. The VNIR sensor will cover the 0.4 - 1.0 micron band with 10 nm spectral resolution (60 bands), and the SWIR sensor will cover the 1.0 - 2.5 micron band with 10 nm spectral resolution (150 bands). Most VNIR and SWIR data will be compressed (statistically redundant spectra removed) on orbit using the ORASIS algorithm. A factor of 10 compression is anticipated. Uncompressed data may be collected occasionally for calibration or other purposes. The uncompressed data cube is about 0.8 GB in size for the VNIR sensor and 2.0 GB for the SWIR sensor. There is also a single band, 5 m pixel resolution at nadir, PIC sensor. The uncompressed PIC data cube is about 0.5 GB in size.

**Level 1B Archived Product:**

The Naval Research Laboratory (NRL) will archive NEMO data. The Level 1B archive product will be in Planetary Data System (PDS) format (probably). When users request Level 1B data they will receive the data collected for the complete 30 km by 200 km scene. A partial list of the contents of the Level 1B archive product is (one each for VNIR, SWIR and PIC sensor):

a. Raw data cube, if collected
b. ORASIS OCP file (compressed image file)
c. PDS Header (summary information for scene)
d. ENVI Header (contains basic data cube information)
e. Calibration Matrix (converts counts to physical units, scaled to 16 bit integer)
f. Wavelength Calibration file
g. Pointing/Geometry File contains information on the location (longitude and latitude) and geometric properties (solar and view zenith and azimuth angles, etc.) of the center and end pixels for each line in the uncompressed image

**Access to NEMO Data:**

The Level 1B archive will be accessible via a Web interface. Using standard browsers, the user will be able to remotely access the NEMO database and search for data using time and/or location. The user can then request specific data sets. Data will probably be delivered on DVD, but the exact media type has not yet been decided upon.

The NEMO Program will provide algorithms for atmospherically correcting the data using ATREM, and for deriving for each water pixel the column averaged values for chlorophyll, Kd(490), CDOM, suspended sediment, and bathymetry/bottom reflectance. Algorithms for reconstructing the data cube from ORASIS compressed files, calibrating the data cube, geo-rectification, mosaicing and COIS/PIC merge
(image sharpening) will also be available. Generally, only executable code will be provided (no source code). However, we are interested in collaborative development of new algorithms and would share source code as appropriate for these efforts. The supported operating systems are Windows NT 4.0 and SGI IRIX. The algorithms can be run using the host operating system command line. There is, however, utility in bundling them within a unifying package. With that in mind, our current baseline is to also bundle our algorithms with the Environment for Visualizing Images (ENVI). ENVI compatible NEMO related software will be provided, but the ENVI application itself must be purchased from RSI. Using ENVI, access to algorithms and data will be available through a point and click interface.

The algorithms will also be hosted at the NRL archive and users will be able to access them via the Web interface. As part of our HyCODE participation, NRL will process NEMO data for HyCODE sites through at least the atmospheric correction algorithm. We are particularly interested in checking our calculated normalized water leaving radiances obtained from NEMO data with those measured by other groups, and seek active participation from other HyCODE investigators to help validate calibration, atmospheric correction and other NEMO algorithms using HyCODE data.

**NEMO Availability and Timeline:**

Launch is currently planned for December 2000. It is anticipated that HyCODE sites will be used in the data verification phase that will take place during the first 90 days after launch. Observations will routinely begin 90 days after launch. NEMO observations of HyCODE sites will be made once per week, on average. NEMO data will be processed to Level 1B format and archived 3 days after downloaded to ground. HyCODE NEMO data will be atmospherically corrected within 2 weeks after being archived, resources permitting.

ii. SeaWiFS data for the HyCODE program

SeaWiFS data are being collected and processed for the LEO-15 and West Florida Shelf regions in real-time. NRL Stennis is currently processing the data and placing them on their web site (web7240.nrlssc.navy.mil/ocolor). Access to SeaWiFS requires a NASA approved investigator. SeaWiFS data are currently being processed for coastal areas using a modified SEADAS (NASA) algorithm. This has been modified to improve the atmospheric correction in coastal waters and to retrieve the following products:

Remote Sensing reflectance $R_s$ (6 channels, 412, 443, 490, 510, 555, 670)
Chlorophyll (SeaBam)
Diffuse attenuation coefficient, $K_{490}$
Diffuse attenuation coefficient, $K_{532}$
Absorption total (6 channels)
Backscattering total (6 channels)
Atmospheric aerosols.

The SeaWiFS data that are placed on the web are in gif files that have been processed from HRPT received SeaWiFS data. These gif files are generated from HDF files of the SeaWiFS level 1a data. These HDF files have been processed to account for calibration, atmospheric correction, coastal iterative processing, and registration to a Mercator projection. The HDF files can be input to the SEADAS NASA software for display and image manipulation.

The level 1 SeaWiFS data are available from the NASA SeaWiFS project to authorized users.
The Ocean Color Section at NRL SSC will be collecting and processing SeaWiFS data and generating a time series of bio-optical products starting in January 1999 through Dec 2000. These gif products will be available in near real-time beginning in summer 1999.

c. Modeling Data Policy (Paul Bissett)

One of the 6.1 goals of the HyCODE program is to facilitate the interactions between numerical experiments and observational experiments, thus leading to an adaptive sampling program that optimizes both the modeling and data collection programs. The data policy described in this document focuses on the in situ collected data. In light of the above goal, it will also be necessary to archive representative simulation experiments for the HyCODE sites. This includes all optical, physical, biological, and chemical numerical experiments.

This policy is not to be interpreted that all of the gigabytes of numerical output will be placed on a central location for access by the entire HyCODE investigative team. Instead, representative output as it is available, in the form of two- and three-dimensional plots should be placed on the data archive servers at each HyCODE site. The policy is meant to be flexible and it is anticipated that the exact form of the output to be archived will be determined via interactions among the participants. It is expected that data and metadata will be delivered to the distributed data site (one of the following as appropriate: LEO-15, WSF, CoBOP) and to the Navy within 6 months of collection. Representative simulation data will be delivered as developed. It was agreed that LEO-15 and WSF data would not be password protected. Note that the following statement will be included on the distributed sites: "The collector of these data have right of authorship when data are used by others."

It should be noted that the archiving of modeling data is not as simple as just running the simulation and storing the output, because the models are still being developed. It is much the same as trying to develop a new instrument and archiving every signal that the instrument puts out without understanding what the signal means.

d. Special Navy Requirements and Transitions (Bob Arnone and Walt McBride)

The products generated from the HyCODE program are scheduled for transition to the Naval Oceanographic Office (SSC). These products will provide important contributions to support various operational warfare centers. These include Mine Warfare, Amphibious Warfare, Special Warfare, and Coastal Warfare and Non-acoustic Anti-Submarine Warfare. These Navy products will provide real-time environmental characterization of optical properties on coastal waters. These products will enhance the swimmer visibility and electro-optical systems (laser line scan, laser imaging systems, etc.) which rely heavily on water clarity and laser propagation. Additionally, HyCODE will enhance Navy systems in coastal areas in bathymetry and high-resolution coastal environmental reconnaissance.

The HyCODE data will be used in the development of new algorithms used to support the development of satellite algorithms and to provide the navy operations a method of characterizing the spatial and temporal variability on bio-optical data. These in situ data will provide the architecture for in situ optical data-basing to be used for NAVO operations.

HyCODE transition to NAVY operations:
1. Real-time bio-optical data from specific NEMO test sites (WFS, LEO-15, and Lee Stocking Island)
2. Data-basing architecture and data format for bio-optical data
3. Algorithms for NEMO (hyperspectral) and SeaWiFS (multi-spectral) data
4. Sample products from the test sites to NAVO.
5. Software procedures and methodologies to NAVO
6. Estimates of performance of EO system performance
7. Estimation of diver visibility from in situ and satellite systems.
8. Establish the procedure for accessing and validating the HyCODE products at NAVO.

**e. The Rutgers LEO-15 RODAN Data System**

The Long-term Ecosystem Observatory (LEO-15) uses a project-driven data management system for observations and data-assimilative model results. Datasets are placed in four categories or levels. Level 1 is the raw data as recorded by specific instruments, either in binary, or often translated into ascii and converted to engineering units. The ascii files are usually generated with software supplied by the instrument manufacturer. These are considered the most raw form of the data for most applications. The raw files are usually stored on tape and eventually transferred to optical media. These datasets usually are managed by the scientists that collected the data.

Level 2 consists of any of the numerous processed data products that are constructed from a single sensor or sensor system. The products are assumed to be quality controlled, and ready for scientific use at this point. These files are usually structured ascii format, since this type has been found to be the most useful for data validation and quality control. The Rutgers Ocean Data Access Network (RODAN) is being developed to manage these datasets.

Level 3 datasets are constructed by compositing various Level 2 products. Examples of this are daily files containing all the available CTD data from numerous instruments, overlays of CTD/Optical data, overlays of CODAR surface currents on satellite SST or ocean color. Again, the usual format is ascii, with some advanced products in netCDF. Data management is via RODAN.

Level 4 are the full 3-d datasets generated by the data-assimilative models. These files are almost exclusively netCDF format. Systems being considered to manage the full 3-d datasets include Ferret.

The Rutgers Ocean Data Access Network is being developed by scientists and programmers at the Institute of Marine and Coastal Sciences. The purpose is to provide simple access to the LEO-15 datasets to a wide range of users in both the scientific community and to the general public. Three main access points are envisioned.

Most users are expected to access the datasets via the World Wide Web, browse through them, and download the data they desire for further processing on their own. A second group of users includes the established DODS scientific community. Since DODS uses a Matlab GUI, interactive access by the general public through DODS is limited. But DODS users have the ability to access the datasets directly through C or Fortran program calls. The third group of users are those building Distributed Information Systems that use CORBA protocols to access datasets from a wide network of computers for assimilation in numerical models. While RODAN is constructed with DODS and CORBA compatibility in mind, our biggest user group prefers to access the data via the World Wide Web. We therefore have concentrated development tasks on providing Web access.

RODAN is operated by a virtual controller (a web server) that has access to several databases and the associated meta-data, analysis modules (usually written in Matlab, C or Fortran), and Visualization software (Matlab or Java Applet). The Web interfaces on the virtual controller for RODAN are written in
Java, C and Perl. When you access the site from the Web, a GUI will appear on your screen that explains what data is available, provides access to the metadata, and asks for information on your specific request (ex. time periods of interest, processes to be run, do you want to plot the data, or download ascii files). The virtual controller will access the data and process it using the available analysis modules. If a plot is requested, the controller will plot the data and send it back to the user for display on their screen. If only datafiles are requested, they can be downloaded and stored on the users computer.

Datasets currently accessible from the Web (http://marine.rutgers.edu/cool/rodan.html) include the Underwater Node Data, the Meteorological Tower Data, the Coastal Ocean raDAR (CODAR) Data, Archived Moored, Towed and AUV ADCP data and Declouded AVHRR Sea Surface Temperature Data.

f. The WOOD System (Jeff Smart)

The long-term goal of the World Ocean Optical Data (WOOD) system is to provide a comprehensive worldwide database containing a broad range of optical data, including the diffuse attenuation coefficient, the beam attenuation coefficient, the backscattering coefficient, and bioluminescence potential. The database is easy to use, internet accessible, and frequently updated with data from new at-sea measurements. The database supports a wide range of applications, such as environmental assessments, sea test planning, tactical utility assessments, and Navy mission planning. Eventually it will allow for extrapolation of optical parameters in lieu of direct measurements (e.g., beam attenuation will be estimated from diffuse attenuation and backscatter data), and an error estimate will be provided for the extrapolated results.

The website is equipped with data display tools (such as online maps and X-Y plots) to allow the user to quickly discern the utility of the available data to meet his or her needs. Below are more details.

WOOD Graphical Mapping Features:
* Interactive world map of coastlines, including:
  -- Zoom in/out, scroll right/left, re-center map
  -- Display of the lat/long at position of the mouse pointer
  -- Display of range between mouse clicks on the map
  -- Selection of region either from map or manual entry
* Display data locations as symbols on the map, including:
  -- Edit symbols & colors used to represent data on the map
  -- Color symbols by season or month
  -- Toggle symbol visibility by parameter, season, or month

WOOD X/Y Plotting Features:
* X/Y Plot all data for a specific data parameter
  -- Uses same colors for season/month as map
  -- Option to plot only data currently visible on map
  -- Option to plot every nth data point
* Highlight data symbol(s) on map & same curve(s) on X/Y plot
  -- Option to plot only highlighted data

In order to provide these online graphics, we have implemented automatically downloaded JAVA applets to display data maps and profiles. Oracle Version 7.3.3 database software is used to store all metadata and optics-related data, and Oracle Webserver software Version 2.1 and the JDBC Java interface software were selected to connect the database to the internet. The entire system is installed on a Pentium PC running under the Windows NT Server operating system. In October 1998, the database computer was
upgraded to a dual processor 333 MHz Pentium using multiple hard disks to distribute the processing load. An exact duplicate system has been purchased to support development of the new capabilities (online graphics, data extrapolation, and relational searches). This approach allows access to the existing operational database system while simultaneously developing new displays and features of the improved system. For more details, please see http://www.jhuapl.edu.

g. Data Formats (Prepared by Jeffrey H. Smart (JHU/APL); inputs provided by Joseph Wielgosz (JHU/APL) & Weilin Hou (USF))

In approaching the question of how to streamline the sharing of HyCODE data, the team assigned to answering this question wanted to allow for maximum flexibility of various data types, while at the same time imposing minimal extra work on those who submit data. In order to provide an uninhibited free flow of HyCODE data among the key users (observationalists, modelers, and algorithm developers), it was decided that a single, well-defined (or "structured") ASCII exchange format must be adopted. This decision came about after lengthy discussions about the pros and cons of various formats, ranging from completely unstructured ASCII formats (i.e., whatever is most convenient for the PIs who collect the data) to very structured, more complex binary formats, such as netCDF, HDF, and PDS. (NetCDF is one of the most popular gridded formats in use by modelers; HDF is similar to netCDF and is used by NASA for SeaWiFS data; PDS is the format selected for NEMO satellite data.)

To the modelers, the advantages of these binary formats are that 1) they are designed to handle evenly spaced imagery data, 2) they are self-documenting (which means that their header information defines all the details about the data structure), and 3) they use much less disk space than ASCII formats. However, observationalists generally will not use these kinds of formats for a number of reasons. First, in contrast to modelers and algorithm developers, their data do not start out being in one of these formats, so an extra step is required to convert to them. Second, in many cases the custom analysis software used observationalists cannot handle these complex binary formats, so they tend to use their own (or commercial instrument defined) formats. Finally, ASCII is nice to use because a human can read it and because all types of computers (Macs, PCs, and Unix systems) can read it.

As a result of these considerations, it was decided that a well-defined structured ASCII exchange format is a reasonable compromise between what the modeler/algorithm developers need and what observationalists use. A great advantage of choosing a single exchange format is that a single conversion program (from the ASCII to say netCDF) can be written. Furthermore, observationalists can share data directly using the ASCII exchange format. Numerous other oceanographic projects have also chosen to share data using a structured ASCII exchange format, and it makes sense to examine their formats before defining our own. For example, NASA’s SEABASS database requires all users to submit their data in a single format of this type. By doing so, they were able to build an automated database loading scheme. The HyCODE project intends to follow a similar approach to handling the large volume of data.

Summary of Proposed Structured ASCII Exchange Format

In designing a structured ASCII exchange format, we have examined existing database and dataset designs used in the World-wide Ocean Optics Database (WOOD), in NASA’s SEABASS bio-optics database, and in NOAA’s Ocean Climate Laboratory (OCL) format. Based on this study, we propose to create an expanded version of the SEABASS format (denoted hereafter as SEABASS-X) that is backward compatible with SEABASS. We propose that SEABASS-X be used by all ONR PIs, especially those involved in the HyCODE and CoBOP programs. The rest of this section explicitly defines the SEABASS-X format.
Proposed List of Allowed Data Types:

SEABASS defines the following data types:

/cast - optical profiles, CTD profiles, etc.
/pigment - same as cast except can only contain pigment data
/drifter - drifting sensor
/mooring - fixed location time series recording
/along_track - moving ship or aircraft, continuous measurements like from AC-9
/above_water - above_water sensors (i.e., Lw, Rrs) such as SIMBAD
/sunphoto - sun photometer (handheld)
/scan - spectrophotometric scans, etc.
/multiple - this defines a file with entries representing multiple stations. An example is a chlorophyll tables where it holds all the data for a cruise.

To this list we need to add the following more general data types to handle random space-time series data, such as those collected by a tow-yo system or by an AUV or UUV:

/tow-yo - data collected on a paravane performing a repeated depth profiling pattern (e.g., sawtooth or sinusoidal)
/4-d - continuously changing location and time data

From the HyCODE database perspective, several of these data types will use common sets of attributes. For example, cast and pigment are both just different types of single lat/lon profiles. These kinds of data only need a start_lat and a start_lon to define their location. Multiple is just a collection of data of type cast, but if we have only a single metadata header, then we need to provide an array of times, latitudes, and longitudes to properly describe this data type. Drifter and along_track are really simple versions of the 4_d data type.

In order to build the relational database using automatic read routines, we need to precisely specify the format for each of these allowed data types. When possible, we will try to minimize the parameters required to define each type. For example, virtually all ocean data types require DEPTH as a parameter, and MET data could use negative "depth" values to indicate height above the water surface.

Before the data exchange design is formally completed, it would be very helpful if PIs would send in samples of the data files they currently generate so we can examine them and verify that our design encompasses all crucial information. We would also greatly appreciate receiving lists of preferred abbreviations to use for variables (e.g., Kd, Ed, Lu, bb, etc.) because an approved list of abbreviations must be generated.

Allowed Attributes:

As indicated in these tables, the REQUIRED metadata attributes have been kept to a minimum. The list of allowed attributes is much longer than the REQUIRED parameters so that 1) additional useful data can be stored as desired, and 2) all the functionality of NASA's SEABASS database is preserved. However, in one important sense we are being more restrictive than SEABASS. The reason is that SEABASS only puts metadata into a relational database; actual data is only kept as an assembly of data files in sorted directories. In contrast, we will be building a relational database of the metadata and the data itself. Therefore, we must be more explicit in defining the allowed data types and their associated attributes. For
example, SEABASS would accept an acronym like "Lu". We will need to identify Lu data and provide its wavelength as a separate attribute, so for optical parameters we intend to enforce the use of acronyms that include wavelength (e.g., Lu488). Also, the SEABASS system never checked the actual data columns themselves, but in our case we need to check the "correctness" of these data. Therefore, a data type like "tow-yo" will be checked to ensure that it includes columns for DEPTH, LAT, LON, and TIME for each sample.

The following list explicitly specifies the allowed metadata attributes:

/begin_header (denotes start of header; this and all subsequent header lines MUST be preceded by a forward slash (/) and the lines must end in carriage return)
/start_date (yyyyymmdd)
/end_date (if differs from /start_date)
/start_time (hhmm, e.g. 1245 or decimal time; e.g., 12:45 is 12.75 )
/end_time (if differs from /start_time)
/time_code (conventions allowed: local or GMT; if local, provide offset; e.g., "LOCAL,-4" for Eastern Daylight Time)
/north_latitude (northernmost latitude; e.g. 19_ 30.0' or decimal location e.g. 19.5000; values south of equator are negative)
/south_latitude (southernmost latitude; if not provided, default value is north_latitude)
/west_longitude (format same as Lat; use -180 to +180, i.e., WEST values negative)
/east_longitude (easternmost latitude; if not provided, default value is /west_longitude)
/data_type1 (e.g., single cast, multiple casts, point sample, tow- yo, etc.; see proposed definitions at end of this section)
/fields1 (e.g., Depth,Temp,Sal,Ed488,Lu532, ...; optical parameters MUST include center wavelength (in nm); all parameters must come from our list of standard abbreviations; we need PIs to contribute to the creation of this list)
/units (provide one string for each /data_type; a standard list/format of approved units will be specified)
/contact (e-mail address of person responsible for these data)
/start_wavelength (in nm; applies only to /data_type=scan data)
/end_wavelength (in nm; applies only to data_type=scan data)
/incremental_wavelength (in nm; applies only to data_type=scan data)
/experiment (e.g. HyCODE or CoBOP)
/cruise (a string like "July 1999 LEO-15," max length = 71 chars)
/station (a string like "Sta005," max length = 71 chars)
/investigators Names of PIs &/or persons providing the data
/affiliations Institutions, Universities, Government agency, etc of PIs
/parameters (generic data types, e.g., "CTD"," IOPS"," AOPS"," currents")
/delimiter (allowed values: comma, space, TAB, semi-colon; default is SPACE)
/missing (place holder for missing data; default is -9999)
/cloud_percent (provide as a percentage, e.g., 50)
/wave_height (significant wave heights,peak-to-trough, in meters)
/wind_speed (in knots or meters/sec)
/wind_unit (knots or meters/sec; required if /wind_speed provided)
/original_file_name (name of source file used by PI)
/calibration_files (name of file(s) used to calibrate the data)
/documents2 (cross-references to other related files)

[the following approved but ALWAYS optional parameters are being added to the SEABASS list]
/bottom_type (e.g., mud, sand, silt, clay, rock, gravel)
/min_sample_depth (shallowest depth of data in this file; depths below surface are positive downward; above surface values must be negative)
/max_sample_depth (deepest depth of data in this file)
/water_depth3 (bottom depth in meters; defaults will be assigned from DBDB-V)
/instruments1 (see example definitions under Metadata at http://wood.jhuapl.edu)
/secchi_depth (meters)
/npts (number of data points in the profile or time series)
/edit_status1 (e.g., 0=raw data, 1=spike edited, etc., see possible definitions under Metadata at http://wood.jhuapl.edu)
/quality1 (e.g., 1=good, 2=fair, 3=poor, 4=unknown, 5=mixed)
/bandwidth (in nm; applies only to optical measurements, e.g., 10 nm for MER1048)
/vertical_resol (in meters; e.g., 10 m if a 10 m linear least squares fit pass was applied)
/dissemination (yyyymmdd, i.e., date that data can be made public via ONR WOOD database)

The last header attribute MUST be the following:
/end_header@ (denotes end of header block)

Attribute Notes:

1. We will maintain a set of allowed values/definitions on a website. In the case of QUALITY, we realize this is a subjective parameter. As an example of how to use it, assume a PI collects Ed profiles to compute Kd. If there is uniform sunshine throughout the cast and the instrument tilt is minimal, Kd quality is probably "good." However, if variable cloudiness exists during the profile, quality is probably "fair" or "poor." If most of the cast is "good" but a limited portion is "poor," then "mixed" quality is assigned. If calibration is in question but the relative structure is deemed "accurate" (as in a fluorometer used to compute chlorophyll or a poorly calibrated c-meter), then a value of "good relative structure" could be used. To aid in assigning a quality value, we suggest using "good" for data having a combined accuracy (due to calibration and noise levels) of better than 10%; use "fair" for a combined accuracy of 10 to 20%; use "poor" for combined accuracy worse than 20%. Avoid using "mixed" if the worst portions have an accuracy of at least 20%; "mixed" should be used on combinations of "good" and "poor" data.

2. For SEABASS, which contains files, this link tells where to find additional documentation. For the HyCODE database, which is not just a flat file system, the same approach would apply except that we will have to precede the file name with the source file's URL. This could be done automatically when the data are extracted from the source FTP site.

3. DBDB-V is currently the best unclassified source for world-wide bathymetry, but in local areas like the LEO-15 site we may use better data such as NOAA's NGDC Coastal Relief Model for the US NE Atlantic Coast. If a user provides WATER DEPTH with their data, then their value will not be replaced with any of these other sources.

Additional Specifications:

To be SEABASS compatible, the values assigned to the header attributes, described above, cannot contain either a backslash (/) or the "@" symbol, and underscores should be used rather than white space. (For example, specify the PI's name as Jeff_Smart rather than Jeff Smart. In the database itself, the underscores will be replaced by white space.) Also, the values associated with each header element (except for /start_header and /end_header@) must be preceded by an equals sign and terminated with a carriage
return. For example, use:

/fields=Ed488,Sal,Temp
/units=uW/nm/m^2,ppt,degC

In general, the International System of Units (SI units) are to be used and exponents are denoted using the caret (^) symbol followed by the exponent in French braces. An approved list of valid FIELDS abbreviations will be provided separately and data providers are requested to send in additional abbreviations as needed.

Comments:

Unlimited additional comments can be provided to further characterize the data, such as "variable cloudiness during downcast; sunny during upcast; HPLC chlorophyll method." Comments will be preceded by an "!" character on each line.

Data Conversion Software:

The HyCODE relational database will be patterned after the World-wide Ocean Optics Database (WOOD) developed and housed at JHU/APL. JHU/APL will also be building this new database, and before loading the data into the HyCODE database, APL intends to convert the ASCII exchange format files into XML (eXtensible Markup Language) ASCII files. XML has numerous advantages:
* it has extensive documentation and existing software libraries
* software already exists to verify "correctness" of files and parse their contents
* it will facilitate automated data loading
* it is supported by recent Netscape and Microsoft browsers

APL intends to provide their ASCII exchange format-to-XML conversion program to PIs. The PIs can use this program to test their exchange format data for "correctness," and if they like they can apply this program to all of their data and just submit the resulting XML data files. Alternatively, APL will run automated routines to convert the data into XML and subsequently perform automated loads into the relational database.

As a service to the PIs, we also intend to provide software to automatically handle a few well-known commercial ASCII formats (e.g. WETVIEW files, RDI/ADCP "T" files, Eric Firing netCDF ADCP format, Sea-Bird CTD files, etc.), and we would have a script available that recognizes these formats and automatically converts them to our exchange format. However, these commercial formats, such as WETVIEW files and Biospherical "PAC" files, do NOT include critical required data like latitude and longitude, so we would still have to force people to provide a separate ASCII file containing those data. (Even the RDI files, which generally include DATE, POSITION, and TIME data, won't have the CONTACT information.)

h. Data Policy Issues (Tommy)

The general responsibilities of HyCODE PIs include:
1) Produce highest quality data possible
2) Document all measurement and analytical techniques (utilize metadata) [HyCODE protocol working groups will provide guidance documents]
3) Estimate and report accuracy and precision for each measurement
4) Follow HyCODE data sharing policy.

Key information which PIs need to supply includes:
1) site description
2) data description
3) common format data files
4) high resolution data on CD-ROMs
5) data for WOOD and possibly other databases and archives

It is expected that data and metadata will be delivered to the distributed data site (one of the following as appropriate: LEO-15, WSF, CoBOP) and to the Navy within 6 months of collection. It was agreed that LEO-15 and WSF data will not be password protected. Note that the following statement will be included on the distributed sites: "The collector of these data have right of authorship when data are used by others."

i. Data Flows and Requests (Paul Bissett)

Figure 7. gives a schematic view of the anticipated data inflows from the PI's to the data archive servers, and the outflows from the data servers to the requesting researcher. It is expected that the originating PI will keep all of the high frequency temporal and spatial data collected by their individual instrumentation. These data are designated as Level 0 data, and with the appropriate quality controls made available to HyCODE participants via direct request to the originating PI. For convenience, the PI's quality-controlled data location is designated as the Local Data Archive Server (LDAS). The originating PI will define the format of this data.

Level 1 data are a temporal and spatial subset of the Level 0 data. It is to be processed to location, depth, and time bins as described earlier. These data are then to be placed in the Structured ASCII format (also described in Section XX), and forwarded to the Distributed Data Archive Server (DDAS). The DDAS will then make these data available via Web and FTP access to the HyCODE community. The Structured ASCII format is to be consistent between similar types of data, i.e. AC-9 from all participants will have the same Structured ASCII format, such that read/write routines to ingest these data into analytical software packages need only be created once. This will ensure rapid access to field data for scientific analysis and quality control by the entire HyCODE community. The Structured ASCII format includes header information that lists the methodology used in the processing of the data, flags the data for potential problems, and give directions for requesting the original Level 0 data.

Level 2 data are a temporal and spatial subset of the Level 1 data. These data will be placed into a structured data format as discussed earlier, and forwarded from the DDAS to the Global Data Archive Server (GDAS). It is not the responsibility of the PI to forward the structured data to the GDAS, as the DDAS administrator will be responsible for this transition. Again, the data header will give the methodology of processing, potential errors, and the appropriate information necessary to locate the originating Level 1 data will be given in a header file.

Data requests from the community can enter the HyCODE data system at three locations, i.e. the LDAS, DDAS, GDAS. This structure is meant to facilitate the transfer of information between PI’s, and not result in an onerous burden upon the originators of data. As such, data requests should flow to the appropriate data archive server that meets with the requesting researchers temporal and spatial needs. Upon deciding that a higher resolution is needed, the header information within the data record should point the user to the appropriate archive site and file location (see Figure 1 ). For example, a data user wishing to look at AC-9 profiles from the Middle Atlantic Bight, might first check with the GDAS for the relevant
information available. After viewing the GDAS data, the user may find that he/she may need a greater temporal and spatial resolution specifically from the LEO-15 site. The user may look to the GDAS record header information for the location of these files, or may request this information directly from the LEO-15 DDAS. Requesting data in this fashion will ensure participation by all investigators, and not create friction between those collecting data and those requesting data.

4. Summary of General Issues

The consensus of the group was that

a) it would be efficient to use the RODAN system for the WFS and CoBOP data management implement and WOOD software would also be utilized for cross-referencing

b) data from HyCODE data management groups for LEO-15, WSF, and CoBOP would port optical data to WOOD and NAVO

c) PIs would store final processed high-resolution data on CDs and submit data following scheme outlined in Data Policy section

d) common protocols for instrumentation would be used (e.g., IOPs, AOPs, discrete data, etc.); information concerning calibration procedures, processing software, formatting would be provided as metadata

e) previews of data (large amount of data warnings) will be provided in distributional databases and updating of data (e.g., new versions) will be monitored

f) example model runs would be made available to HyCODE PIs.

5. Timeline of Actions

April 27, 1999 5pm PST: Input material to Tommy DONE

May 6, 1999: Draft Version 1.0 back to group for input and modification DONE

May 12, 1999: Feedback to Tommy for final revisions DONE

June 14, 1999: email to HyCODE PIs about Workshop and its recommendations/Place Report on Web.

July 1999: RODAN Test at LEO-15

July 1999: HyCODE PIs welcome to visit LEO-15

November(+/-) 1999: HyCODE General Meeting at Rutgers including Data Management subgroup

December 1999: LEO-15 accepting data to be put on RODAN

6. Assignment of Tasks
a) Observational Timelines:
LEO-15 (Oscar Schofield) DONE/NEED MAP OF LEO-15 FIELD SITE
WFS (Jeff Donovan) DONE
CoBOP/LSI (Weilin Hou) NEED?
NEMO Avail. and Timeline (Algorithm, data format, amount, ...) (Bill Snyder) DONE

b) Table of Measurements by Site (Parameters, sampling rates/vertical resolution, etc., telemetry or not, wavelengths of light measured, instruments, Responsible PI):
LEO-15 (Oscar Schofield) DONE
WFS (Jeff Donovan) DONE
CoBOP/LSI (Weilin Hou) DONE

c) Figure Showing Overview of Data Flow (Paul Bissett) DONE

d) Formatting (Jeff Smart) DONE

e) RODAN Description (Scott Glenn)

f) WOOD Description (Jeff Smart) DONE

g) Modeling Data (Paul Bissett and Scott Glenn) DONE

h) Data Policy Statement (Tommy) DONE

Table Captions
Table 1. Data available, observational platforms, and instrumentation for LEO-15. (See Attachment: LEO)
Table 2. Data available, observational platforms, and instrumentation for west Florida shelf (WFS). JEFF, CAN YOU SEND THIS AS AN ATTACHMENT AGAIN? SORRY!
Table 3. Data available, observational platforms, and instrumentation for CoBOP/Lee Stocking Island (LSI). (See Attachment: hycode-cobop.doc and hycode-cobop.pdf).
Table 4. Required combinations of data type and header attributes.

Figure Captions
Figure 1. Field site map of LEO-15.
Figure 2. Timeline for LEO-15 activities. (See Attachment: HyCODE- timeline.PDF)
Figure 3. Field site map of WFS. (See Attachment: current-shelf.pdf or current-shelf.eps)
Figure 4. Timeline for WFS activities. (See Attachment: HyCODEdatasheet.doc or HyCODEdatasheet.pdf).
Figure 5. Field site map of CoBOP/LSI.
Figure 6. Timeline for CoBOP/LSI activities. WEILIN, I NEED THIS FIGURE. THANX!

Figure 7. Schematic view of the anticipated data inflows from the PI’s to the data archive servers, and the outflows from the data servers to the requesting researcher. (See Attachment: Modeling Data Policy.doc)

### Appendix I Attendees

<table>
<thead>
<tr>
<th>NAME</th>
<th>AFFIL.</th>
<th>E-MAIL/ADDRESS</th>
<th>PHONE/WEBSITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steve Ackleson</td>
<td>ONR</td>
<td><a href="mailto:ackless@onr.navy.mil">ackless@onr.navy.mil</a>, ONR 321, Ocean Optics, ONR, Arlington, VA 22217-5660</td>
<td>(703) 696-4732</td>
</tr>
<tr>
<td>Bob Arnone</td>
<td>NRL</td>
<td><a href="mailto:arnone@nrlssc.navy.mil">arnone@nrlssc.navy.mil</a>, Code 7343, Code 7343, Stennis Space Center, MS, 39529</td>
<td>(228) 688-5268, web7240.nrlssc.navy.mil/ocolor</td>
</tr>
<tr>
<td>Paul Bissett</td>
<td>FERI</td>
<td><a href="mailto:pbissett@marine.usf.edu">pbissett@marine.usf.edu</a>, Florida Environ. Res. Inst., 701 Channel Side Dr., The Florida Aquarium, Tampa, FL 33602</td>
<td>(813) 273-4072</td>
</tr>
<tr>
<td>Tommy Dickey</td>
<td>UCSB</td>
<td><a href="mailto:tommy@ices.ucsb.edu">tommy@ices.ucsb.edu</a>, Ocean Physics Laboratory, Univ. of California, Santa Barbara 6487 Calle Real, Unit A, Goleta, CA 93117</td>
<td>(805) 893-7354, <a href="http://www.ices.ucsb.edu/opl/opl.html">http://www.ices.ucsb.edu/opl/opl.html</a></td>
</tr>
<tr>
<td>Jeff Donovan</td>
<td>USF</td>
<td><a href="mailto:jeff@marine.usf.edu">jeff@marine.usf.edu</a>, MSD, USF, 1407th Ave. South, St. Petersburg, FL 33701</td>
<td>(727) 553-1116</td>
</tr>
<tr>
<td>Scott Glenn</td>
<td>IMCS</td>
<td><a href="mailto:glenn@carribean.rutgers.edu">glenn@carribean.rutgers.edu</a>, Rutgers, New Brunswick, NJ 08525</td>
<td>(732) 932-6555 x544</td>
</tr>
<tr>
<td>Weilin Hou</td>
<td>USF</td>
<td><a href="mailto:hou@monty.marinunsf.edu">hou@monty.marinunsf.edu</a>, MSD, USF, 1407th Ave. South, St. Petersburg, FL 33701</td>
<td>(727) 553-3988</td>
</tr>
<tr>
<td>Walt McBride</td>
<td>PSI</td>
<td><a href="mailto:wmcbridge@nrlssc.navy.mil">wmcbridge@nrlssc.navy.mil</a>, Bldg. 9121, Stennis Space Center, MS 39529</td>
<td>(228) 689-8458</td>
</tr>
<tr>
<td>Jeff Smart</td>
<td>JHU/ APL</td>
<td><a href="mailto:jeffsmart@jhuapl.edu">jeffsmart@jhuapl.edu</a>, Rm. 8-308, JHU/APL, Johns Hopkins Rd., Laurel, MD 20723-6099</td>
<td>(443) 778-4331, <a href="http://wood.jhuapl.edu">http://wood.jhuapl.edu</a></td>
</tr>
<tr>
<td>Oscar Schofield</td>
<td>Rutgers</td>
<td><a href="mailto:oscar@ahab.rutgers.edu">oscar@ahab.rutgers.edu</a>, Rutgers, New Brunswick, NJ 08525</td>
<td>(732) 932-6555 x548</td>
</tr>
<tr>
<td>William Snyder</td>
<td>NRL</td>
<td><a href="mailto:snyder@rsd.nrl.navy.mil">snyder@rsd.nrl.navy.mil</a>, Code 7212WS, Naval Research Lab, Washington DC 20305</td>
<td>(202) 404-1376</td>
</tr>
</tbody>
</table>

Cool Web Sites
Appendix II: Workshop Agenda

April 12

9:00am Welcome to workshop/Local arrangements and logistics (Scott and Oscar)
9:15 Discussion of objectives and goals of Workshop and any revisions to agenda (Tommy/ONR Program Managers)
9:45 Discussion of the general types of data HyCODE will be collecting Observations: LEO-15, WFS, LSI (Reps. of each project)
10:30 Coffee Break
11:00 Discussion of satellite/aircraft algorithms/data (Bill Snyder/Bob Amonen/Brian Scheiber)
11:30 Discussion of modeling "data" (Paul Bissett)

12:00 Lunch

1:30 Special Navy requirements/transitions (ONR Program managers/Walt McBride)
2:00 Discussion of schemes for organizing and transferring data sets (e.g., How web can be used, what is key information for web (e.g., timelines, data, model examples, etc.)? (Jeff Smart and others)
3:00 Tour/navigation of Rutgers data management system (RODAN) (Rutgers group)
5:30 Adjourn for day

April 13

9:00am Review first day discussions/select critical topics for day’s discussions (Group)
9:30 Data policy issues (Tommy/ONR Program managers)
10:15 Coffee Break
10:45 Selected topics (e.g., computer resources, network requirements how much documentation/metadata/calibration, etc.) (Group)

12:00 Lunch

1:30 Discussion of assignments of responsibilities for various aspects of HyCODE management (Group)
2:30 Plans for organizing/distributing/utilizing data sets (Group)
3:15 Break
3:45 Other relevant topics (Group)
4:15 Discussion of workshop report (Tommy)
5:00 Adjourn workshop

Appendix III David Porter Message

On our CMO page you will find that we use that format on all our projects. You don’t want to spend time reinventing HTML. So here are some thoughts:
1) Re-use HTML code as often as you can:
   a) Set up a common format for all the investigators
   b) Share all your knowledge
   c) Maybe use a common style sheet

2) You can use Frames if you want.

3) Decide on one site to be the focal point (like our CMO page). It will merely link to the other sites and the other investigators.

4) Have a table that gives all the investigators, their emails, their addresses and phone numbers.

5) The individual sites should have:
   a) FTP area. This can be password controlled by the end institute.
   b) Any written documents (science planning, questionnaires, etc. should be in html and rtf format. The rtf format allows you to download the document into your word processor.

6) As far as images, moving icons, fancy backgrounds go I would implore all people to follow the rule set up by Mr. Disney in making movies. He said if the song or the scene does not move the story line along then it should not be there. If a rotating globe in the upper right corner of a page doesn’t help in understanding what the author is trying to get across, it should not be there.

Feel free to use our CMO code as an example. The idea is to distribute the data (and hence the work) to the people who have an interest in the data. The less passwords the better.

**Tables**