SURA
The Science of Collaborative Research

A Community Coastal and Ocean Modeling Testbed to Improve Understanding and Operational Forecasts of Extreme Events and Chronic Environmental Conditions Affecting the U.S.

Principal Investigator: Richard Luettich, University of North Carolina, Chapel Hill
rick_luettich@unc.edu
Program Manager: Sara Madden, SURA

Award Number: NA13NOS012013 ***Award Period:*** September 1, 2013 - August 31, 2018

Coastal and Ocean Modeling Testbed Data Management Plan
Updated February 5, 2018

Southeastern Universities Research Association
Washington, D.C.

Authors:
C. Reid Nichols[1], Richard A Luettich, Jr.[2], L. Donelson Wright[1], Marjorie Friedrichs[3], Alexander Kurapov[4], Andre J. van der Westhuysen[5], Katja Fennel[6], Eoin Howlett[7], Linda Akli[1] and Gary Crane[1]

[1] Southeastern Universities Research Association (SURA), Washington, D.C.
[2] University of North Carolina, Chapel Hill, North Carolina
[3] Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA
[4] Oregon State University, Corvallis, OR
[5] Environmental Modeling Center, National Oceanic and Atmospheric Administration
[6] Dalhousie University, Halifax, Nova Scotia, Canada
[7] RPS ASA, South Kingstown, RI

I. Introduction

The US IOOS Coastal and Ocean Modeling Testbed (COMT) exists to accelerate the transition of scientific and technical advances from the coastal and ocean modeling research community into improved modeling products and services for operational use. Improvements may include model results that are more timely, more reliable, more accurate, show decreased uncertainty over time or are less expensive (in computational costs or financial) than legacy models. The Southeastern Universities Research Association (SURA), which manages the COMT, relies on stakeholder feedback to help assess project performance and to adapt projects and activities to maximize the value of their outcomes (Figure 1). SURA's approach to managing the COMT has included the facilitation of High Performance Computing access, transitioning models, model improvements, best practices, tools and other capabilities into operational use to improve our understanding and predictions of the consequences of extreme events and chronic conditions on the coastal waters of the United States. Projects are intended to result in innovative products that include new or improved model code, new knowledge regarding model operations, archived data for future model evaluation, innovative software tools, and lessons learned.

Figure 1. SURA's approach to managing the COMT has involved iterative engagement of stakeholders, modelers and management and has included the facilitation of High Performance Computing access.

The COMT began in 2010 and results from the first three years were reported in a special issue of the *Journal of Geophysical Research* in late 2013; a general overview of the goals and results of this initial phase can be found in Luettich *et al.* (2013). The current COMT includes five projects, each of which is working to transition models, model improvements, best practices, tools and other capabilities into operational use to improve our understanding and predictions of the consequences of extreme events and chronic conditions on the coastal waters of the United States. The current projects are (1) Chesapeake Bay Ecological Forecasting; (2) Integration of US West Coast Operational Coastal and Ocean Models; (3) Puerto Rico/US Virgin Islands Inundation and Wave Forecasting; (4) Gulf of Mexico Ecological Forecasting; and (5) Improving the Supporting COMT Cyber Infrastructure.

The U.S. IOOS Blueprint (U.S. IOOS, 2010) names three functional subsystems as crucial to a fully coordinated enterprise: (1) observations and data transmission; (2) modeling and analyses (the prime COMT activity); and (3) data management and communication (DMAC). The COMT data management has been designed to support IOOS DMAC principles and standards and to ensure that project execution, data interpretation and curation follow smoothly from the research tasks. Activities that comprise the data management component of the COMT have and will continue to evolve over time, relying on expertise, requirements, and tools developed by organizations such as the National Oceanic and Atmospheric Administration (NOAA), the National Science Foundation (NSF), and the National Aeronautics and Space Administration (e.g., Cornillon, et al., 2003, Cornillon, et al., 2009, Gallagher and Milkowski, 1995). The COMT supports current NOAA and NSF policies on the dissemination and sharing of research results and adhere to the following best practices (e.g., NOAA, 2013, 2015 and National Science Foundation, 2016):

1) prepare and promptly submit for publication, with authorship that accurately reflects the contributions of those involved, all significant findings from work;
2) share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work;
3) share software created under COMT awards. Make software and the ensuing products widely available and usable; and
4) make results, data and collections available to the scientific community, while observing the principal legal rights to intellectual property.

This document outlines the current status of the COMT data management structure and procedures. The COMT team anticipates that the plan will continue to evolve as projects mature, key data sets requiring curation are identified, and data and metadata collection methods are refined. Thus we expect the plan will be updated annually.

II. COMT Data Management Statement of Purpose

Open communication and data sharing are critical to the success of the COMT. Therefore, the primary data management goals are to ensure the fidelity and accessibility of data among COMT participants; to minimize the amount of time COMT participants need to spend on data management activities to create high quality data and metadata; to make COMT project data available to the COMT cyber-infrastructure team for their used in tool development and testing; and to ensure that important data and metadata can be discovered and utilized by the broader scientific community. To achieve these goals, the following have been identified as principal priorities for the COMT data management:

1) the efficient storage and sharing of large datasets by and between COMT project participants;
2) the definition and adoption of data formats and metadata standards to encourage the development of analysis and visualization tools that are able to access data from multiple sources and types;
3) the archival and curation of key datasets from COMT projects for future use by the coastal and ocean modeling community at large including the use of a Digital Object Identifier (DOI®) System to identify COMT data sets (Coordinate with University Data Librarians);
4) the advertisement and access of archived data sets from the COMT to the coastal and ocean modeling community at large via the US IOOS COMT website: https://ioos.us/comt and other data outlets; and
5) the migration of key datasets to federal data archives, such as the National Centers for

Environmental Information (NCEI), for long term data archival.

III. COMT Data Repository

The COMT data repository is currently housed on two servers in the Department of Physics and Astronomy at Louisiana State University (LSU). The repository hardware is managed by Ms. Hortensia Valdes who has more than 20 years of experience managing research data with the LSU Physics and Astronomy Department. Ms. Valdes monitors storage capacity, runs backups, installs system software updates and ensures the servers and data are protected from security threats. The repository's current capacity is approximately 20 terabytes; a little more than 50 percent of which is presently in use. Network-attached storage (NAS) was added during 2016. NAS is specialized for serving files either by its hardware, software, or configuration. An important benefit of NAS is its ability to provide multiple clients on the network with access to the same files. See Appendix A for detailed hardware specifications of the COMT data repository.

During the execution of COMT projects, the repository is available for participants to house data sets and share these data with their collaborators. Examples include model inputs (e.g., forcing, grids), model outputs and companion observational data. Prior to project completion, data that are of archival quality will be identified by the COMT data management team. These will be retained following the project completion and made available to the broader coastal and ocean modeling community who may be interested in using them for future modeling studies (either as a part of or independent from future COMT activities). The COMT cyber-infrastructure project team will implement catalogs that are searchable by automated software including search engines, other catalogs, and decision-support software. These catalogs will provide the ability to run various kinds of searches and get metadata records or summarization of metadata in the results. The cyber-infrastructure team is also using the COMT repository as the source of data for the development of visualization and analysis tools such as the COMT model viewer, which facilitates the display of COMT data, the comparison of model results with observations and cross-model evaluation.

As required, selected data will be transferred or linked to the US IOOS web portal, http://comt.ioos.us/map/ and to appropriate national data centers (e.g., National Centers for Environmental Information or NCEI). Provision of the data to a national data center will release SURA from long-term (>10 year) data preservation. SURA will discuss the operation and maintenance of the COMT data repository and the longevity of data contained therein as part of project closing with the U.S. IOOS Program Office.

The repository is available to all COMT project participants and the broader modeling community via the COMT testbed viewer (see http://oceansmap.com/comt/). The repository will provide data access via multiple open source protocols including OPeNDAP via THREDDS, which are in use by US IOOS. COMT participants are also able to access the repository via other standard remote access protocols (e.g., Secure Shell or SSH and File Transfer Protocol or FTP).

The status of the COMT data repository will be reported as part of SURA's semi-annual COMT progress report to NOAA IOOS.

IV. COMT Data Management Team

Successful COMT data management is a collateral responsibility requiring central oversight and assistance and engaged participation by the COMT project teams. Thus COMT data management requires an effective data management team as outlined below.

Ms. Hortensia Valdes
COMT Data Manager – Data management team leader
SURA
Louisiana State University
Department of Physics & Astronomy
Baton Rouge, LA

Ms. Sara Madden
COMT Program Manager
SURA
Washington, DC

Dr. Rick Luettich
COMT lead-PI
University of North Carolina at Chapel Hill
Institute of Marine Sciences
Morehead City, NC

Dr. Marjorie A.M. Friedrichs
Chesapeake Bay Ecological Forecasting
Virginia Institute of Marine Science, College of William & Mary,
Gloucester Point, VA

Dr. Alexander Kurapov
Integration of US West Coast Operational Coastal and Ocean
Models Oregon State University
Corvallis, OR

Dr. Andre van der Westhuysen
Puerto Rico/US Virgin Islands Inundation and Wave
Forecasting
Environmental Modeling Center
National Weather Service
Silver Spring, MD

Dr. Katja Fennel
Northern Gulf of Mexico Ecological Forecasting
Department of Oceanography
Dalhousie University
Halifax, NS, Canada

Mr. Eoin Howlett
Cyber Infrastructure
RPS ASA,
South Kingstown, RI

To lead the data management effort, SURA has designated a COMT Data Manager with the following responsibilities:

1) leads the COMT data management team;
2) works with the data management team to define and document COMT data and metadata standards;
3) facilitates the use of the COMT data repository by project participants, reviewing data in the repository with the project PIs on at least a semi-annual basis;
4) ensures that all data sets identified for archival are complete and conform to COMT standards, e.g., formatting, providence and metadata;
5) acts as a liaison between the COMT modeling teams and the COMT cyber-infrastructure team to facilitate the development of tools for use with data contained in the COMT data repository such as the COMT model viewer;
6) transitions selected archival data to permanent storage at an appropriate national data center;
7) reviews and helps maintain the US IOOS COMT website https://ioos.us/comtdescribing research activities and identifying available data sets for the benefit of the broader coastal and ocean modeling community;
8) liaises with COMT project teams and U.S. IOOS DMAC personnel on data management issues;
9) coordinates data needs with the manager of the COMT data repository at LSU and reports on hardware status, disc space capacity and usage, etc. for inclusion in the COMT semi-annual report; and
10) leads discussions on data management at future annual PI meetings.

V. Types of COMT Data

The current COMT modeling projects cover the Chesapeake Bay, Caribbean Sea, Gulf of Mexico, and the Pacific Ocean along the US West Coast. Participants are (i) compiling in situ observational data and supporting numerical model data to define model forcings, for data assimilation and for model skill assessment and (ii) generating input data for and output data from the numerical models they are using. Data represent both physical variables (e.g., meteorological properties, water levels, currents, hydrographic properties) and ecological variables (e.g., nutrient concentrations, chlorophyll concentrations, oxygen levels). These data are disparate in their sources, structure and formats as described briefly below and in further detail in Appendix B. When possible COMT will use established IOOS DMAC protocols or other national / international standard protocols for the management of these data. If unavailable, protocols will be developed and documented as part of the COMT data management program.

Observational data may come from many sources including instruments operated by NOAA, (e.g., National Data Buoy Center buoys, the National Water Level Observation Network, and the Chesapeake Bay Interpretive Buoy System), IOOS (e.g., HF Radar and gliders), and the USGS (e.g., USGS Surge, Wave, and Tide Hydrodynamics (SWaTH) Network, river gauging stations), and ship based hydrographic cruises (run by many different agencies). Variables represent either scalars or vectors (Appendix B) and data types include geospatial fields, single point time series and combinations of both in raster, vector, text and Network Common Data Form (NetCDF) formats. Supporting numerical model data come from large scale atmospheric or oceanic models, typically from NOAA / NWS or the US Navy. Data are usually in gridded fields in either GRIdded Binary (GRIB) or NetCDF formats. The specifics of the grid (both horizontally and vertically) vary depending on the numerical model. These data are typically used to provide surface and lateral boundary conditions for COMT models.

Models currently being used in the COMT include the Advanced Circulation model (ADCIRC), the Finite-Volume, primitive equation Community Ocean Model (FVCOM), the Navy Coastal Ocean Model (NCOM), the Sea, Lake, and Overland Surges from Hurricanes model (SLOSH), Simulating Waves Nearshore (SWAN), WaveWatch III® (WWIII), the Regional Ocean Modeling System (ROMS) and several ecological models. Model inputs include bathymetry; meteorological fields; river discharge; and scalar properties / fluxes along open boundaries. Model outputs include temperature, salinity, velocity, sea surface height, wave spectra, significant wave height, wave period, wave direction, dissolved oxygen, etc, on both structured and unstructured grids and may be in the form of 2D or 3D scalar and vector fields and single point time series. Wave spectra are typically comprised of energy as a function of frequency and propagation direction, either on a 2D grid or at point locations.

Storage requirements for the above described data are considerable given requirements to store and serve interim data during project execution and to archive and serve key data sets following project completion. The resources currently available in the COMT data repository are described in Section III.

VI. COMT Data and Metadata Standards

Metadata are required to trace data origins and interpret the contents of a data file. The sources of most COMT metadata are the COMT project participants who are collecting observations, running models, etc. The COMT data repository anticipates receiving a range of file formats (e.g., NetCDF, ASCII, Excel) for both data and metadata (see examples in Table 1). One of the initial tasks of the COMT Data Manager will be to reach consensus with IOOS DMAC representatives and the COMT data management team on COMT data and metadata standards. Data and metadata file formats and standards are expected to evolve throughout the COMT project and should provide clear information regarding the data sources.

DOIs can be created for data sets and are equivalent with DOIs for other scholarly publications. They enable accurate data citation, and most importantly provide online access to COMT data for discovery, attribution and reuse. DOIs are a persistent identifier and as such carry expectations of curation, persistent access and rich metadata. Metadata for DOIs may include names, identifiers, descriptions, types, classifications, locations, times, measurements, relationships and any other kind of information related to COMT datasets.

Table 1. Example data and metadata file formats for each COMT task area.

Task Area	**Known hData File Format**	**Known Metadata file format**	**Known Metadata Standards**
Bottom Characteristics	Bitmap image (BMP), **Bathymetric Attributed Grid (BAG), Geospatial Portable Document Format (GeoImage),** MultiResolution Seamless Image Database (**MrSID), Tagged Image File Format (TIFF)**	XML SGML	FGDC
Physical Oceanography	ASCII	SGML	FGDC
Ecology	ASCII	SGML	FGDC
Model Input	ASCII, NetCDF		
Model Output	NetCDF, GRIB, BUFR	SGML	FGDC, NetCDF CF Metadata Conventions, ISO 19915, NCML

ASCII - American Standard Code for Information Interchange
BUFR – Binary Universal Form
CF – Climate and Forecast
FGDC - Federal Geographic Data Committee
GRIB – Gridded Binary
NCML – NetCDF Markup Language
SGML - Standard Generalized Markup Language
XML – eXtensible Markup Language

The range of metadata required to make the collected information useful to scientists, data centers, and the public varies greatly within COMT. Identifying the metadata necessary to scientifically describe the observations, and model inputs and outputs will be accomplished by the COMT Data Manager working with the remainder of the data management team. The goal is to provide an optimal amount of metadata so that the data is fully described but future users are not overwhelmed. While metadata collection/generation is often labor intensive, we will endeavor to reduce the workload on the scientists by applying forms or other tools to aid the metadata collection/generation. In some cases, the COMT researchers are unfamiliar with metadata formats, so the Data Manager will work with the scientists to develop mutually beneficial solutions. Prior to archival, data and metadata will be evaluated to ensure that they meet COMT standards and / or the standards of national data centers such as NCEI.

Table 2 lists the types of metadata that scientists from each project have determined to be useful for their research activities. This list will evolve as the project develops. Whenever possible, we will utilize existing metadata standards (see Table 1); however, many of the researchers are not familiar with standards, ontologies, or vocabularies. We anticipate that some cross-walking of individual science party vocabularies to national standard vocabularies will be required and this will be accomplished by the COMT data management team. The nature of the research being conducted is anticipated to result in ancillary documentation. These documents may include event logs, data processing software, model code, etc. When these are submitted to the COMT data center they will be required to include metadata necessary to trace them to a specific COMT funded activity.

Table 2. Preliminary list of metadata that project teams have identified as important for the scientific application of their data and analyses.

Task Area	**Metadata**
Bottom Characteristics	Date, area, equipment used, name of survey platform; the geodetic reference system used, i.e. horizontal and vertical datum including ties to a geodetic reference frame based on ITRS (e.g. WGS84) if a local datum is used; calibration procedures and results; sound speed correction method; tidal datum and reduction; uncertainties achieved and the respective confidence levels; any special or exceptional circumstances; rules and mechanisms
Physical Oceanography	Time (UTC), latitude, longitude, depth, SI units, data parameters
Ecology	Station number, date, time of sampling, depth of sample, latitude, longitude, SI units, quality controlled, data parameters
Model Input	Climate and Forecast metadata standards
Model Output	Spatial parameters, time, iteration counters, tolerance limits, who ran model, model used, version # (note we had a model metadata template that we used in COMT 1)
Metadata are used to provide documentation for data products. In essence, metadata answer who, what, when, where, why, and how about every facet of the data that are being documented.	

VII. Policies for Access and Sharing and Provisions for Appropriate Protection/Privacy

Data files and model output will be housed at the COMT data repository (Section III). All uploaded datasets are cataloged within the COMT THREDDS Data Server (TDS) using NcML to aggregate in both time and across variables as necessary. TDS gives users several options to access the datasets including HTTP, OPeNDAP and OGC standards for geospatial data such as WMS, SOS and CSW. The geospatial web services enable COMT to operate a map server which has been used to develop an interactive model viewer for each project. The capabilities of this map server will be augmented to allow display of COMT data from buoy data, gliders, hindcasts, etc. The map server connects directly to the TDS and services using the TDS to provide the user a range of options to view and download COMT data and model output. The details of the database and map server structure will be developed over the coming year.

Permission to access the COMT data will occur at three levels:

1) the investigator collecting/creating the observation or model output,
2) among COMT collaborators and partners, and
3) the general public.

The first two levels are available while a project is underway. Upon completion of a COMT project, key data sets will be moved into the archive and opened to the community.

As we proceed with development of the database to track individual data sets and model output, mechanisms will be put in place to provide release authorization for each of these levels. We anticipate that many data sets will be made available among COMT partners and that public release will correspond to the completion of the COMT project. Basic discovery metadata must be provided for all data sets to be released to the public. All metadata information is stored alongside the unmodified datasets and will be made available via web services to all interested parties as soon as possible after receipt of the metadata from the individual investigators. The Data Manager will also establish a data submission protocol that will include a requirement to capture essential metadata for each data set submitted to the center.

COMT will comply with federal guidelines including the U.S. Department of Commerce (DoC) policy on Electronic Transmission of Personally Identifiable Information (PII) (ref. U.S. Office of Management and Budget, Memorandum M-07-16, “Safeguarding Against and Responding to the Breach of Personally Identifiable Information”). Per DoC policy as published by the Office of the Chief Information Officer, PII is defined as “information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother’s maiden name, etc.” PII is further classified as “Sensitive” or “Non-Sensitive”. “Sensitive” PII includes:

1) Social Security Numbers (including truncated to last 4 digits)
2) Place of birth
3) Date of birth
4) Mother's maiden name
5) Biometric information

6) Medical information, except brief references to absences from work

It is the responsibility of the individual researchers to remove any PII prior to submission and any PII located by COMT data center staff will be deleted upon receipt. Note, for reference, the COMT data center will accept “Non- Sensitive” PII which may include the following:

1) Work, home, and mobile phone numbers
2) Work and home addresses
3) Work and personal e-mail addresses
4) Resumes that do not include an SSN or where the SSN is redacted
5) General background information e.g., Biographies
6) Position descriptions and performance plans without ratings

VIII. Policies and Provisions for Re-use, Redistribution

As noted above key data deposited in the COMT data repository will be archived and made available to the broader coastal and ocean modeling community after a project is completed. The range of future users of the data and model output from the COMT research is unknown. Since all future uses of the data cannot be foreseen, the COMT data and metadata standards will strive to cover the range of metadata required to describe the data long after the conclusion of the project. Submission of the data to the national data centers will preserve the legacy of the project for the foreseeable future.

IX. SURA Facilitated HPC Access

Until completion of the current contract, SURA Information Technology staff will continue to facilitate access to national high performance computing resources through NSF Extreme Science and Engineering Discovery Environment (XSEDE) for computationally intensive model runs. To date, SURA has acquired nearly 10 million service units on HPC assets for COMT through XSEDE and has worked with testbed participants to port codes and resolve problems on these systems During the period from 2016 through 2017, the usage of XSEDE resources by Katja Fennel and Marjorie Friedrichs has increased from 10 million and 10.75 million. Typically, time is allocated for one-year and the data storage resources are only available while the project is active. Likewise, these resources do not provide support for public or shared interactive access to data. Thus, for both local and national computational and data visualization resources other resources are necessary for the long- term archival and preservation of the resulting data products.

X. References

Cornillon, Peter, James Gallagher and Tom Sgouros, 2003. OPeNDAP: Accessing Data in a Distributed, Heterogeneous Environment, *Data Science Journal*, Vol. 2, pp. 164-174.

Cornillon, Peter, Jennifer Adams, M. Benno Blumenthal, Eric Chassignet, Ethan Davis, Steve Hankin, James Kinter, Roy Mendelssohn, James T. Potemra, Ashwanth Srinivasan, and Joseph Sirott. 2009. NVODS and the development of OPeNDAP. *Oceanography*, Vol. 22, No. 2, pp. 116–127.

Federal Geographic Data Committee. Available online. URL: https://www.fgdc.gov/. Accessed January 26, 2016.

Federation of Earth Science Information Partners. Available online. URL: http://www.esipfed.org. Accessed January 26, 2016.

Gallagher, James and George Milkowski, 1995. Data Transport within the Distributed Oceanographic Data System, *World Wide Web Journal*, Issue 1, pp. 691-702; Conference Proceedings, Fourth International World Wide Web Conference, December 11-14, 1995, Boston, Massachusetts, Newton, MA: O'Reilly and Associates, 1996.

National Science Foundation, (2016, January 25). Proposals and Award Policies and Procedures Guide, NSF 16-1, OMB Control Number 3145-0058. Available online. URL: http://www.nsf.gov/pubs/policydocs/pappguide/nsf16001/gpg_print.pdf. Accessed February 14, 2016.

NOAA, 2015. Data Sharing Directive for NOAA Grants, Cooperative Agreements, and Contracts, Version 3, Available online. URL: https://nosc.noaa.gov/EDMC/documents/Data_Sharing_Directive_v3.0.pdf. Accessed February 4, 2016.

NOAA, 2013. NOAA Environmental Data Management Framework, Available online. URL:https://nosc.noaa.gov/EDMC/framework.php. Accessed February 4, 2016.

NOAA National Operational Model Archive & Distribution System. Available online. URL: http://nomads.ncdc.noaa.gov/. Accessed February 4, 2016.

NOAA North American Mesoscale Forecast System (NAM). Available online. URL: https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-mesoscale-forecast-system-nam. Accessed February 4, 2016.

NOAA Testbeds & Proving Grounds, National Oceanic and Atmospheric Administration, Available online. URL: http://www.testbeds.noaa.gov/. Accessed January 26, 2015.

OPeNDAP, Open-source Project for a Network Data Access Protocol. Available online. URL: http://www.opendap.org. Accessed January 26, 2016.

Open Geospatial Consortium. Available online. URL: http://www.opengeospatial.org/. Accessed January 26, 2016.

Sgouros, Tom, James Gallagher, and Nathan Potter, 2006. OPeNDAP Server Installation Guide, Revision 14459, Available online. URL: http://www.opendap.org/pdf/install.pdf. Accessed January, 26, 2016.

Appendix A — COMT Server and Data Storage Hardware Specifications

SURA will provide two COMT servers – The original server contains the observed data and provides a backup for the second server. SURA provides COMT with a PowerEdge R720 rack server and a Redundant Array of Independent Disks (RAID) 5, with 16 x 16GB Registered DIMM (Dual In-line Memory Module). The RAID 5's distributed parity evens out the stress of a dedicated parity disk among all RAID members. Read performance is increased since all RAID members participate in serving of the read requests. If a single drive were to fail, subsequent reads can be calculated from the distributed parity such that no data are lost. While the registered memory is more expensive because of the circuitry required, it meets the scalability and robustness needs for the COMT server. The M-class GPU Computing Modules provide parallel computing processors for high performance computing, especially climate, ecological, and weather modeling. The basic components of the servers are as follows:

-PowerEdge R720 (225-2133)
-Risers with up to 6, x8 PCIe Slots + 1, x16 PCIe Slot (331-4440)
-vFlash SD Slot Filler (318-2036)
-iDRAC7 Express (331-3481)
-NVIDIA TESLA M2075 GPU Computing Module PCIe (320-3673)
-GPU Installation Kit (320-2883)
-Broadcom 57800 2x10Gb DA/SFP+ + 2x1Gb BT Network Daughter Card (430-4428)
-2.5" Chassis with up to 16 Hard Drives (317-8474)
-Bezel (318-1375)
-Power Saving Dell Active Power Controller (330-5116)
-RAID 5 for H710P/H710/H310 (3-16 HDDs) (331-4382)
-PERC H710P Integrated RAID Controller, 1GB NV Cache (342-3531)
-Intel Xeon E5-2670 2.60GHz, 20M Cache, 8.0GT/s QPI, Turbo, 8C, 115W, Max Mem --1600MHz (317-9596)
- Heat Sink for PowerEdge R720 and R720xd (331-4508)
- DIMM Blanks for Systems with 2 Processors (317-8688)
- Intel Xeon E5-2670 2.60GHz, 20M Cache, 8.0GT/s QPI, Turbo, 8C, 115W (317-9610)
- Heat Sink for PowerEdge R720 and R720xd (331-4508)
-16GB RDIMM, 1600 MT/s, Standard Volt, Dual Rank, x4 Data Width (317-9640) - Quantity 16
- 1600 MHz RDIMMS (331-4424)
- Performance Optimized (331-4428)
- 900GB 10K RPM SA SCSI 6Gbps 2.5in Hotplug Hard Drive (342-2971) - Quantity 16
- CD-ROM or DVD-ROM Drive: DVD+/-RW, SATA, INTERNAL (313-9090)
- CD-ROM or DVD-ROM Drive: ReadyRails Sliding Rails with Cable Management Arm (331-4433)
- CD-ROM or DVD-ROM Drive: Dual, Hot-plug, Redundant Power Supply (1+1), 1100W (331-4607)
- CD-ROM or DVD-ROM Drive: Power Cord, C13 to C14, PDU Style, 12 Amps, 2 meter, Qty 1 (330- 3151) - Quantity 2
- CD-ROM or DVD-ROM Drive: No Operating System (420-6320)
- CD-ROM or DVD-ROM Drive: No Media Required (421-5736)

The basic components for network attached storage are:

Synology Disk Station 8-Bay (Diskless) Network Attached Storage (NAS) (DS1815+)

- Quad Core CPU with AES-NI Hardware Encryption Engine
- Four Gigabit LAN Ports with Failover and Link Aggregation Support
- Scale up to 18 drives with Synology DX513
- Expandable RAM Module (Up to 6GB)
- VMware, Citrix, Microsoft Hyper-V Certified
- High Availability and SSD Read/Write Cache Support
- Running on Synology Disk Station Manager (DSM)

Western Digital - WD Red 6TB NAS Desktop Hard Disk Drive - Intellipower SATA 6 Gb/s 64MB Cache 3.5 Inch - WD60EFRX

- Specifically designed for use in NAS systems with up to 8 bays
- Designed for the unique system requirements of NAS environments
- Tested for 24x7 reliability
- 3-year manufacturer limited warranty

Appendix B — Additional information describing COMT Data and Variables

This appendix presents an expansion of Section V COMT Data Types in the data management plan. Participants will adhere to conventions for CF (Climate and Forecast) metadata that are found online at URL: http://cfconventions.org/.

An initial list of variables for inclusion in COMT data standards is as follows:

Air Temperature
Average Wave Period
Barometric Pressure
Chlorophyll-A
Current Direction
Current Speed
Depth
Dissolved Oxygen
Dominant Wave Period
Harmful Algae Blooms
Hypoxia
Inundation
Mean Wave Direction Nitrate
Peak Wave Direction
Peak Wave Period
Primary Swell Direction
Primary Swell Period
Primary Swell Wave Height
Reflectance
Relative Humidity Salinity
Sea Surface Temperature
Secondary Swell Wave Height
Significant Wave Height
Swell Height
Swell Period
Turbidity
Water Level
Wave Period
Wave Steepness
Wind Gust (maximum 5-seconds)
Wind Speed
Wind Direction
Wind Wave Height
Wind Wave Period

Given the disparate nature of COMT data, we have found it useful to develop a simple classification scheme as described below:

1) Historical Data. Data types will include raster, vector, and text files. Raster and vector files

are commonly used to better understand the geography of the study area and to develop bathymetric grids. (This must include information on both vertical and horizontal datums.) Data from hydrographic surveys, meteorological stations, water levels stations, buoys, etc, are largely in a text format and may be available from data centers such as NCEI. Text-based, wave hindcast data may be obtained from the U.S. Army Corps of Engineers. Model grids may be developed from atmospheric and oceanographic data, which were produced by numerical models or derived from observational data. Grids may contain any data that can be represented in a two-dimensional matrix.

2) Observational Data. Investigators will be using sustained measurements of important parameters and phenomena occurring in the coastal ocean. For this reason, data from observing systems such as the Gulf of Mexico Coastal Ocean Observing System and the Mid-Atlantic Coastal Ocean Observing System are critical. Data are obtained from weather stations, moored buoys, drifting buoys, gliders, ship-based and space-based observational networks and related telecommunications facilities. Data are used for model inputs, skill assessments, assimilation, and inter-comparisons.

3) Ecological Data. Ecological data are those observations that directly support elements of NOAA’s Ecological Forecasting Roadmap. Examples include hypoxia and Harmful Algal Blooms (HABs). To observe the drivers of ecosystem change, water column observations include a range of oceanographic measurements to support ecological research. Examples include chlorophyll, Colored Dissolved Organic Matter, dissolved oxygen, fluorescence, nitrate, salinity, temperature, and turbidity measurements.

4) Modeling Data. The COMT projects are generating numerical model output from differing types of model simulations. Models to be used include ADCIRC, FVCOM, NCOM, SLOSH, SWAN, WWIII and ROMS. Forecasts and simulations will include fields of temperature, salinity, velocity, sea surface height, significant wave height, wave period, wave direction, mixed layer depth, dissolved oxygen, etc, on both structured and unstructured grids.

Model simulations focus on key estuarine and coastal ocean study regions (e.g., Chesapeake Bay, Caribbean, California Current, and Northern Gulf of Mexico). General input data for circulation models includes surface and lateral boundary conditions for momentum and tracer-type variables (temperature, salinity, etc). Forcing data may be point or gridded data time series. Output from circulation models will consist of 2D and 3D scalar and vector fields. The general input for spectral wave models includes forcing by wind fields and currents. The wave model output is a description of the wave spectra in frequency and directional space plus aggregated parameters such as significant wave height and the period and propagation direction of the dominant wave.