NERACOOS DMAC Plan

November 22, 2022

1 Introduction

The Integrated Ocean Observing System (IOOS) is a national network of Regional Associations (RA) managed by the NOAA IOOS Program all working to provide ocean observing data and information to a wide variety of end users. As the RA for the Northeast, NERACOOS has a primary goal to produce, aggregate and provide access to ocean observing data from the region. NERACOOS maintains a robust Data Management And Cyberinfrastructure(DMAC) program to support this function.

The IOOS Program Office and key data management representatives from all of the RAs work closely to develop, test and implement standards and recommendations for managing and distributing data. NERACOOS' DMAC team, led by the Gulf of Maine Research Institute (GMRI), has participated in these efforts since the outset of IOOS. The NERACOOS DMAC team has developed and continues to evolve a robust and efficient data management system that leverages the suite of standards and best practices adopted and recommended by IOOS and the DMAC community to provide a centralized source for regional data access. NERACOOS received its initial Regional Information Coordination Entity (RICE) certification from the IOOS program in 2018.

2 DMAC Management Structure

The NERACOOS DMAC management structure includes the NERACOOS Executive Director, NERACOOS Product and Engagement Manager, GMRI DMAC Lead, GMRI Data Manager, and GMRI Web Developer. The NERACOOS Executive Director (Jake Kritzer) provides strategic guidance and oversight for DMAC activities. NERACOOS Product and Engagement Manager (Tom Shyka) works directly with the GMRI DMAC team, provides oversight for DMAC activities, coordinates communications between NERACOOS and the GMRI DMAC lead, and helps coordinate activities of the funded NERACOOS data providers.

The GMRI DMAC Team is led by co-PI Riley Young Morse who manages all NERACOOS DMAC activities and oversees the GMRI Data Manager and Web Developer. Ms Young Morse, DMAC Lead, is the main POC with the DMAC staff from the U.S. IOOS Program Office and with the NERACOOS funded data providers. Additionally, the DMAC Lead contributes to proposal development and NERACOOS data management reporting requirements. The DMAC Lead is the main point of contact for all technical data-related questions and is an expert in the design and development of data products. The DMAC Lead submits quarterly reports to NERACOOS Executive Director for DMAC operations and website services. The DMAC lead also participates in a regular performance review conducted by NERACOOS (See section 1.2.4 of the NERACOOS SOP).

The GMRI Data Manager (Dylan Pugh) oversees the NERACOOS data management system; the ingestion of data from funded data providers and other data sources, the management of data within the data system, data and system back-up, development and maintenance of data access services. The Data manager participates in IOOS data standards development and testing activities and supports NERACOOS funded data providers to ensure that IOOS recommended standards are implemented and maintained.

The GMRI DMAC Lead and Manager maintains regular communication with the U.S. IOOS Program Office through in-person meetings, phone calls, webinars, and emails. Additionally, The GMRI DMAC lead and Data Manager often consults outside DMAC expertise from our sister RAs (or RICEs) and other federal partners with DMAC expertise. This frequent communication ensures that NERACOOS is aware of all new practices and protocols, as promulgated by the IOOS Program Office, and understands how to implement them. The DMAC lead, data manager, NERACOOS staff and the funded data providers participate in regular DMAC calls to discuss new protocols and evaluate implementation of protocols recommended by the U.S. IOOS office.

Whenever the U.S. IOOS Program provides guidance on data management protocols, NERACOOS will respond within 1 month with an assessment of the relevance of such guidance to our DMAC procedures and if appropriate, an estimate of the time it will take us, given resources and capacity, to reach compliance. Once the data management lead receives the recommended protocol, they will take the necessary steps towards its implementation in a reasonable and timely manner. Implementation of new services is only limited by personnel time and expertise. If feasible, NERACOOS will implement new protocols in one year.

The GMRI DMAC team (Team Lead, Data Manager, Web Developer) are responsible for all aspects of the NERACOOS Data Management System as described above as well as the operations and maintenance of the NERACOOS website and products and services provided through the website. The GMRI DMAC team works directly with NERACOOS funded data providers to support and ensure the implementation of data standards and the maintenance of data feeds. The GMRI DMAC team works closely with the NERACOOS Products and Engagement Manager and NERACOOS stakeholders to ensure that products and services meet well-defined end user requirements.

3 Data Sources

The NERACOOS Data Management System (DMS) aggregates observational data streams from providers distributed throughout the region. Provider types are generally organized into NERACOOS funded (see Section 6 for details), federal or state assets, and private research assets and are outlined in Appendix I: NERACOOS Data Source Table. It should be noted that the classification is based on the management of the assets, as a mix of funding from the sources mentioned above supports some assets. Observational data are organized into type (real-time, near real-time and historic) and are described below. NERACOOS also integrates model data from funded partners and federal sources. Model data are not fully described in the data source assessment but are listed in the data source table.

The NERACOOS DMS was deployed in 2009 leveraging the data management processes and data access products developed by the Gulf of Maine Ocean Observing System (GoMOOS) between 2000-2008. The initial system was optimized
for the structure of the deep water oceanographic buoys from the University of Maine (UMaine). These complex platforms have multiple sensors at multiple depths with differing data collection times. Additionally, new platforms and sensors have been deployed to collect bio-geochemical measurements. The data system has evolved to make it more efficient to integrate data from new platforms and sensors from both NERACOOS funded and non-funded data providers in the region.

3.1 Observational Data

3.1.1 Real-time data

Real-time data are largely provided through buoys and sensors that are remotely deployed and autonomously report on regular intervals. These data are typically telemetered from the sensors directly to the data provider where automated QC is applied and real-time data files are produced and distributed. From there, the real-time data streams are then ingested and processed by NERACOOS DMS at regular intervals ranging from 10 minutes to hourly.

3.1.2 Near real-time data

Near real-time data are data that have a lag in availability, usually due to processing or preliminary analysis. The data that fall into this category include MODIS and AVHRR satellite imagery data of the Northeast region. These data are processed daily by UMaine and delivered as satellite imagery by UMaine on a several day lag. The daily and monthly summary climatology data from the NERACOOS buoys are also processed on a lag and are made available the following day (daily files) or the first day of the next month (monthly files).

3.1.3 Historical data

Historical data include the full time series from buoy deployments. As buoys and sensors are recovered during redeployment, the data are processed and updated for QA/QC, then provided to NERACOOS as the official historical data record for the deployment.

3.1.4 Model output

Model output include forecast products from federal as well as NERACOOS funded partners. These output are accessed at regular intervals as they are available and are typically integrated into several products available through the data products as time series output as well as imagery.

3.2 Provider Type

3.2.1 NERACOOS Funded

NERACOOS fully and partially funds the collection of a significant amount of data in the region. The funded data streams undergo the full data management process and must meet basic standards requirements for metadata, file format and QA/QC. The NERACOOS DMAC team works closely with data providers to ensure they have the tools and guidance needed to produce regular, quality real time and historic data. Details of the NERACOOS funded data providers processes are provided in Section 6.

3.2.2 Federal and State

NERACOOS also ingests and displays data from sources not funded by NERACOOS but readily available and relevant to regional partners and end users. These include assets from federal and state providers. While these data are not required by NERACOOS DMAC to meet the requirements of our funded data streams, most of the federal data are made available through standard web services. These data undergo QA/QC procedures required by the Federal or State agency that collects the data. Providers of data in this category include NOAA NDBC, NOAA NEFSC, NOAA CO-OPS, NOAA NERRS, CDIP and USGS.

3.2.3 Private Research

Data in this category may include data sets from academic institutions, private industry, and NGOs. These data are not always continuous and generally aren't available in standard services like NERACOOS or federal assets. They can also tend to be shorter-term deployments due to limited project funding. Examples of these data include WHOI's ESP harmful algal bloom data, University of Maine EPSCoR buoys, wind energy platforms (Mayflower Buoy), and regional water level gauges. NERACOOS endeavors to make these data available to the region and works with the individual researchers to accommodate data as time and funding allow. Most of these data are ingested into our ERDDAP data system and are not included in data products.

Additional detail about the various NERACOOS data sources can be found in the Data Source Table (Appendix I).

4 Quality Control

Delivering high quality data is a key tenant of the NERACOOS mission. The quality assurance procedures that NERACOOS and its funded data providers have implemented help ensure that data delivered by NERACOOS is as accurate and precise as possible.

As described above in the Data Source section, NERACOOS ingests data from various federal agencies and these data streams have undergone QA conducted by those federal agencies. NERACOOS also funds a significant amount of data collection by several non-federal partners, which are primarily academic institutions. The Principal Investigators (PI) and the teams that are selected to operate sensors for NERACOOS are recognized experts in oceanographic data collection. NERACOOS funding agreements require that funded data providers implement QA procedures on all data streams and require that the relevant QARTOD tests be implemented on real-time data streams. The NERACOOS PIs have made concerted efforts to implement QARTOD in near real-time data streams and most are fully compliant (UMaine's 7 platforms, UNH's 3 platforms) and others are still in process (UConn's 3 platforms). The DMAC team is currently working with NDBC and data providers to migrate from FTP to ERDDAP protocols for delivering data to NDBC and GTS which requires implementation of QARTOD (estimation completion by late January 2023). Additionally, the DMAC team is developing processes to implement QARTOD on water level gauges (Charybdis' 3 tide gauges) and expected to be complete in early 2023.

4.1 Sensor Calibration and Operation

The proper calibration and operation of sensors is a key component to data quality assurance. The NERACOOS funded data partners follow manufacturer's operational procedures to ensure proper calibration and operation of sensors.

4.2 Data Provider QA Processes

NERACOOS funded data providers are responsible for conducting quality assurance on data that is delivered to NERACOOS. The quality assurance procedures implemented by the funded data partners are described for each data provider in Section 6. The majority of funded partner data streams undergo a robust QA process that includes many of the required QARTOD tests as well as other QA tests and reviews. All non-federal sourced data served by NERACOOS undergoes at least a minimum level of QA.

4.3 QARTOD

NERACOOS and its funded data partners are in the process of implementing required QARTOD tests on real-time data streams. All data providers have started implementing required QARTOD tests and are working on full implementation of required tests on all real-time data streams. Work is underway to ensure real-time datasets are fully compliant with QARTOD requirements needed for NDBC to access data via ERDDAP, expected to be completed by early 2023. NERACOOS anticipates that implementation of the required QARTOD tests for all data providers will be completed by Summer 2023.

5 Data Management System

The IOOS Program Office and key data management representatives from all of the regional associations work closely to develop, test and implement standards and recommendations for managing and distributing data. NERACOOS' DMAC team, led by GMRI, has participated in these efforts since the outset of IOOS. The NERACOOS DMAC team has developed and continues to evolve a robust and efficient data management system that leverages data standards adopted and recommended by IOOS and the DMAC community to provide a centralized source for regional data access.

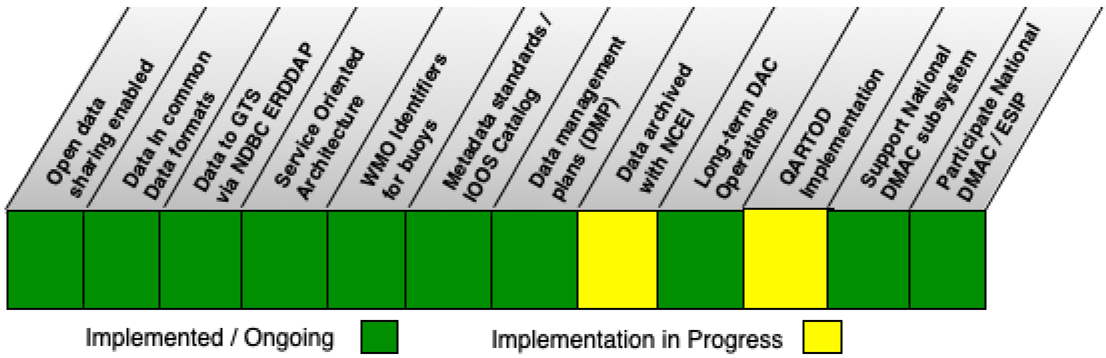
The NERACOOS Data Management System (NERACOOS DMS) is a framework for aggregation, interoperability, discovery and dissemination of observing data (gridded and observational) from the region and is the backbone of the NERACOOS Data Assembly Center (DAC). The GMRI team leverages over twenty years of open-source software development and data management expertise directly on IOOS initiatives in the region. The resulting NERACOOS DMS is based on the suite of standards and best practices developed, tested, and recommended by the IOOS Program Office and the greater IOOS RA DMAC community. The NERACOOS DMS provides the mechanisms and protocols for the full data lifecycle by integrating, aggregating, and distributing data through tools accessible through a centralized access portal (NERACOOS.org). Following the IOOS DMAC Guiding Principles, the DMS provides the core capacity that connects and integrates observations and forecasts, making quality-controlled data discoverable and accessible to stakeholders through a wide variety of information products. The NERACOOS DMAC team continues to work closely with the IOOS Program Office and RA DMAC community to implement new and updated requirements and protocols developed to further enhance access to data.

Aggregating and providing central access to data streams from the NERACOOS funded sensors and models was the primary consideration in the development of the NERACOOS DMS. The DMAC team works closely with NERACOOS data providers to ensure that the highest QA/QC data (historical and real-time) are available in standardized formats for all partners and accessible for products and services through the NERACOOS Data Portal. A secondary goal is leveraging the NERACOOS DMS to include external data streams from state, federal and other independent providers (e.g. academic institutions, private entities, etc.) and make data accessible through the same products and services. Interest in adding new data streams has been steadily on the rise and the NERACOOS DMS has enabled this to be a much more efficient process than in the past.

The NERACOOS DMS is entirely cloud based, deployed on an Amazon Web Services Virtual Private Cloud environment composed of dedicated EC2 instances, S3 storage, and an infrastructure of managed containerized services. The system is backed up nightly, preserving the data, software and configuration of the instances and services for guaranteed disaster recovery. NERACOOS has used a cloud-based system since 2014. The flexibility of a cloud-based infrastructure has made managing the system more efficient by saving costs and making it easier to scale the system and upgrade hardware and software as the NDS evolves.

The NERACOOS DMAC team has been able to implement, manage, and evolve the NDS, following best practices and fully or partially adopt most of the DMAC Subsystem Guiding Principles put forward by NOAA and IOOS (Figure 1).

Figure 1: Status of Implementation of DMAC Subsystem Guiding Principles. Green = fully adopted, yellow = partially adopted.



This section describes the NERACOOS DMS in detail and is organized into the following sections:

1. System Architecture
2. Data Requirements
3. Data Access and Services
4. Hardware and Software Configuration

5.1 System Architecture

The NERACOOS DMS consists of four major components that manage the acquisition, integration, discovery, and dissemination of data from the region. Using interoperability best practices and IOOS recommended standards; data are acquired and ingested through a variety of paths to accommodate data providers with differing levels of technology capacity and to maintain legacy data flow during the conversion to the new system. Once ingested into the NERACOOS DMS, data are stored in one of several ways based on data types. The data are immediately made available via standards-based services for access.

The primary method for acquiring data from data providers is directly from standards-based storage systems through automated protocols using web services and APIs (application programming interface). The majority of NERACOOS data providers have implemented standards-based data storage platforms (e.g., THREDDS: Thematic Real-time Environmental Distributed Data Services, or ERDDAP) to make data available in common formats (e.g., NetCDF). Using this approach data providers are able to store and make available the highest quality data, reducing redundancy and replication and providing direct access to end-users. The NERACOOS DMAC team works closely with the data providers to transform observation data and metadata into compliant formats that conform to data and quality control standards and protocols. As a result, the data managed and curated through the NERACOOS DMS are more interoperable and can be aggregated in region-wide products though the data are served from distributed systems. NERACOOS datasets have been registered with the IOOS Catalog since its inception and are also exposed to the catalog crawler through search-engine friendly Web Accessible Folders (WAF). For data providers lacking bandwidth or capacity to serve data reliably or in compliant formats, the NERACOOS DMS has additional capacity to ingest, apply standards, store, and serve these data through the NERACOOS system. The data access services available through the NERACOOS DMS enables end users to access data directly and programmatically. The protocols and services of the NERACOOS DMS provides a roadmap to integrate new data providers quickly and efficiently (Figure 2).

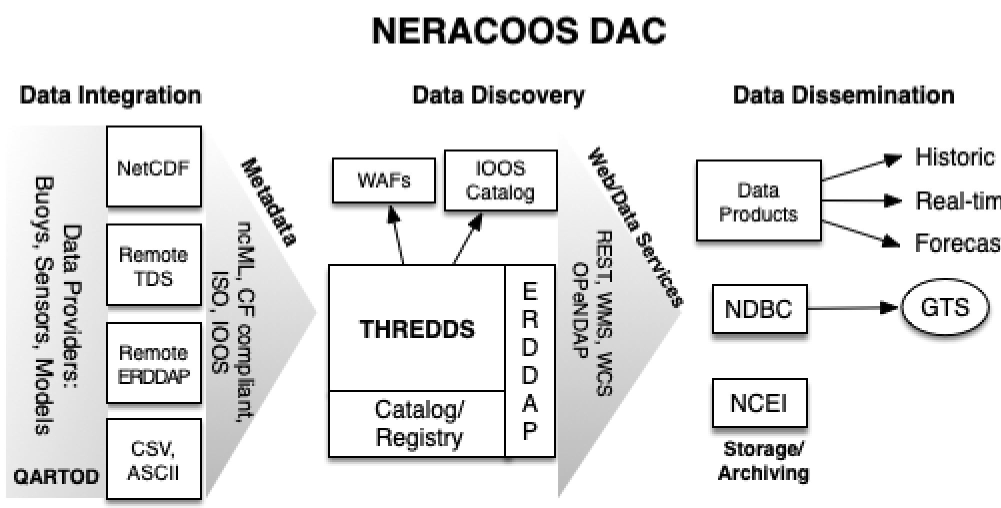


Figure 2: NERACOOS DAC schematic. The data managed and curated through the DMS are interoperable and aggregated in region-wide products, though the data are served from distributed systems. The protocols and services of the DMS provide a roadmap to integrate new data providers quickly and efficiently.

5.2 Data Requirements

Many NERACOOS funded partners have the capacity to store and serve data in standard compliant formats (e.g. NetCDF) that can be ingested by the NERACOOS DMS using automated protocols and services. Guidance is provided to NERACOOS data providers so they can output their data in compliant formats to avoid duplication from multiple versions where not necessary. For data providers lacking bandwidth or capacity to serve data reliably or in compliant formats, the NERACOOS DMS has the capacity to harvest, store and serve these data in standardized output formats.

The basic requirement for funded data providers to contribute data to the NERACOOS DMS is that the data are available in a common format like NetCDF and are compliant with IOOS and CF metadata standards. For data providers not currently serving NetCDF, data can be transformed to NetCDF via scripts or integrated directly into ERDDAP. Metadata are reviewed with the IOOS Compliance Checker and data providers are given guidance on updating records with missing, incomplete, or inaccurate information.

5.2.1 Data Types

NetCDF serves as the basic data format supported for data output. Most observation data providers already use NetCDF as their data storage format, or are capable of producing NetCDF files. Compliance with the latest releases of the Climate Forecast (CF) convention (NetCDF CF 1.6 to 1.8) is required either in the NetCDF headers or via TDS ncML plugin. The protocol to update files with ncML has been well documented and is typically set up once per provider. Updates aren't necessary unless major changes happen with the data collection protocols. Middleware is available to enable data providers to update metadata through ncML (Signell 2014). Additionally, Unidata's Common Data Model, in particular the Discrete Sampling Geometries for Point Time Series, was an important resource in mapping data from the formats used by data providers to be compliant with CF conventions.

5.2.2 Metadata

The NERACOOS framework has and will continue to leverage existing work at national level with regard to conforming data with metadata and vocabulary standards. Data conform to the ISO 19115 and IOOS 1.2 Metadata Standard. The data available through NERACOOS are cataloged locally and kept up to date in a Web Accessible Folder that is accessible to the IOOS catalog crawler.

Additional vocabularies and standards to describe data utilize OGC SWE (SOS, SensorML metadata), CF, MMI, and discrete sampling geometries described in the UCAR NetCDF Common Data Model and Java Library (https://www.unidata.ucar.edu/software/netcdf-java/).

5.3 Data Access and Services

Data accessible from the NERACOOS DMS through machine-machine services and human readable products and tools described in Section 7. The data access and services through NERACOOS utilize IOOS recommended and Open Geospatial Consortium standards. Figure 3 illustrates the wide array of data sources, ingest services, storage and access services that are available through the NERACOOS DMS. Specific details about how each NERACOOS data provider flows data to NERACOOS DMS are outlined in section 6.

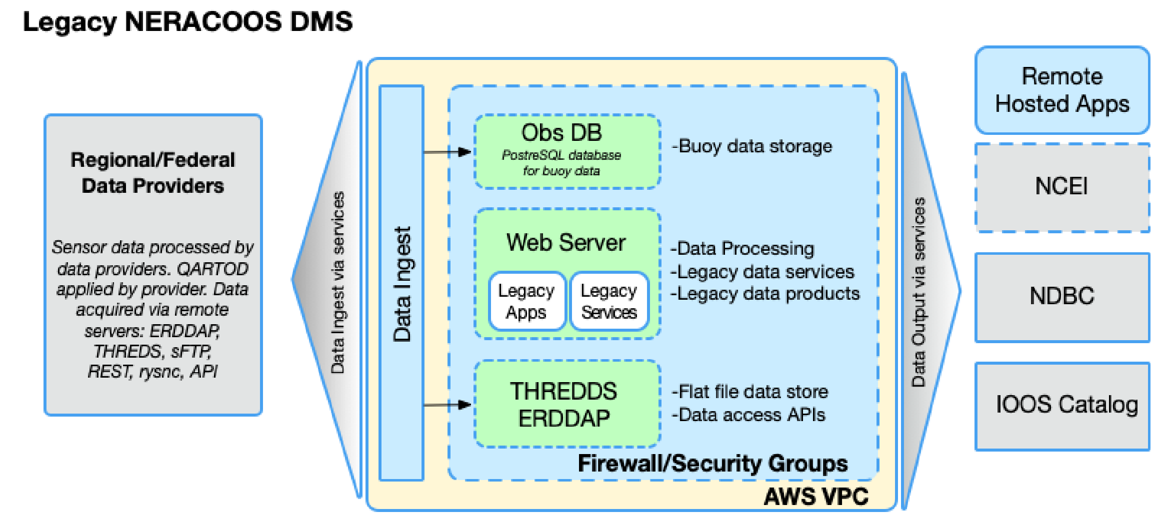


Figure 3a: Legacy NERACOOS Data Management System Schematic

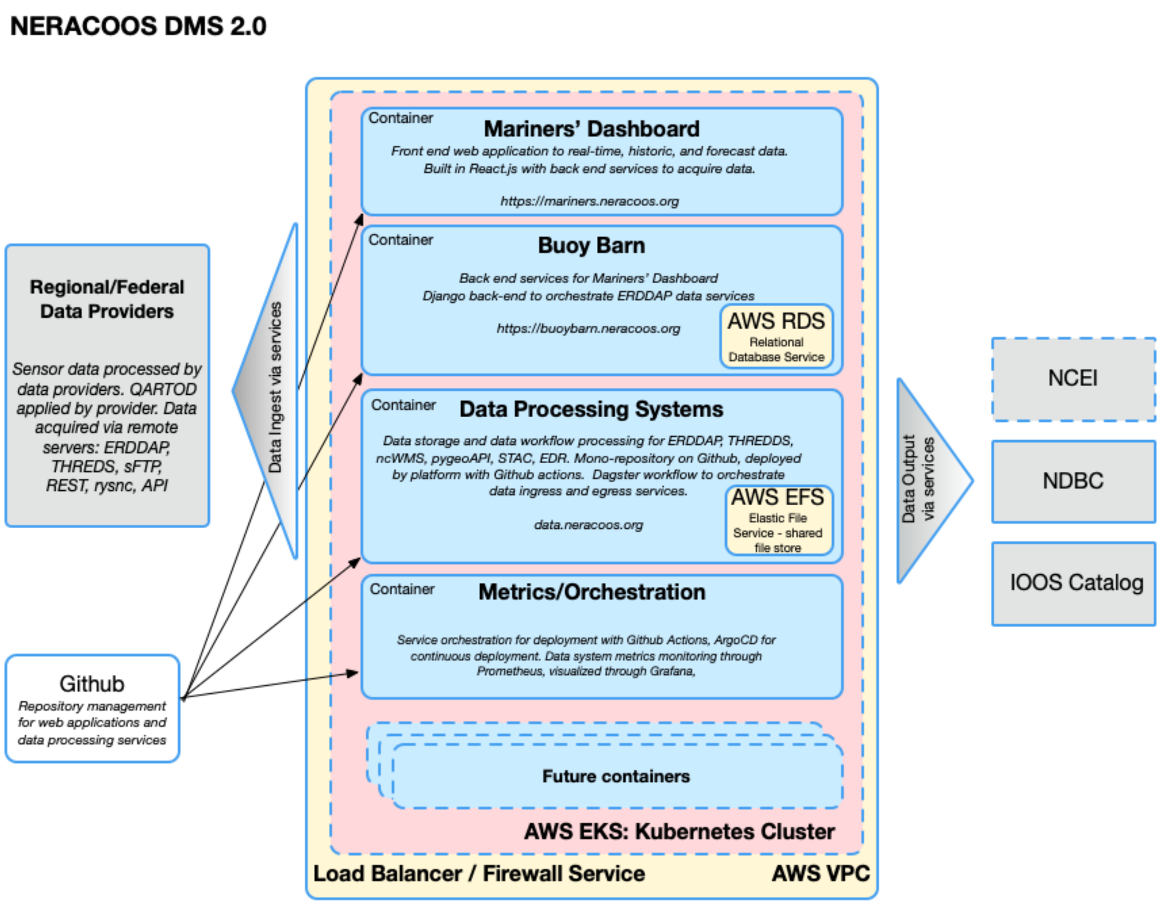


Figure 3b: NERACOOS 2.0 Data Management System Schematic

5.3.1 THREDDS (TDS) Thematic Real-Time Environmental Distributed Data Services

A THREDDS data Server (TDS) is installed on the NERACOOS cloud environment. Data are integrated into TDS via direct TDS to TDS communication with data partners hosting their own remote TDS (UMass Dartmouth, URI, UNH) or by accessing NetCDF files via remote web-accessible directories (UMaine, UConn, BIO) or as local files on stored on NERACOOS servers. The TDS plugin architecture utilizes a suite of standards that handle in-situ observations and models with similar protocols.

TDS is utilized to serve the NERACOOS provider data and metadata in IOOS DMAC compliant formats (ISO 19115, WMS, SOS, OPeNDAP, or subsets of NetCDF). This distributed approach gives data providers control over the data and metadata and avoids replication. Post-recovery files that have been QA/QCd can then replace provisional real-time as they become available. The configuration of TDS/ERDDAP enables it to recalibrate for new data automatically. A new SOS service is created for single parameter requests or time-series requests for data products. TDS can make collections of single files (i.e. all deployments of one buoy that may be in individual files). Various TDS plugins have been installed to accomplish this (ncML, ncISO, ncWMS, and OPeNDAP)

5.3.2 OPeNDAP - Open source Project for a Network Data Access Protocol

OPeNDAP protocols are implemented within the THREDDS Data System.

5.3.3 ERDDAP - Environmental Research Division's Data Access Program (NOAA NMFS SWFSC)

An ERDDAP server, a human readable alternative to TDS, is used as front end for speedier data access and provides output in a variety of usable formats. ERDDAP uses NetCDF as a data source, but other formats such as .csv have been processed successfully for small, discrete historic data files. ERDDAP can serve as a front-end tool for advanced users, but is also an effective backend providing data as a service for products (e.g., Mariners' Dashboard) and other data access tools. ERDDAP produces output files in common outputs: .html table, ESRI .asc and .csv, Google Earth .kml, OPeNDAP binary, .mat, .nc, ODV .txt, .csv, .tsv, .json, and .xhtml. NERACOOS maintains a production ERDDAP server primarily for near real-time data and additional ERDDAP installations for historical collections of ocean acidification and biological data.

5.4 Hardware and Software Configuration

The cloud-based legacy NERACOOS DMS has provided the data infrastructure for the last nine years. While innovative in its inception as a fully cloud-based system, the infrastructure has become increasingly harder to maintain and secure. The NERACOOS DMAC team has developed a migration and upgrade plan for a modernized infrastructure to improve efficiency, security, and scalability. The core functionality of the new NERACOOS DMS (NCDMS 2.0) will follow the high level approach to data management of the current system. Legacy systems, processes, and products will continue to run in parallel with the new system until they are fully replaced with updated infrastructure and products, at which point they will be retired.

Traditional application development, such as that of the legacy NERACOOS DMS, utilizes an approach of deploying multiple software applications and services on specifically configured individual machines. The machines typically run one operating system and maintain the software library for all applications. When individual applications or operating systems aren't regularly upgraded, or reach the point where they can't be upgraded, they create dependencies on older versions of software that limit the ability to upgrade operating systems and software libraries. These issues can make the entire system vulnerable to downtime failures and increased security threats.

The legacy NERACOOS DMS is contained on three discrete EC2 server instances located within a private AWS Virtual Private Cloud. The EC2 instances were selected and optimized for the particular service they would provide. These include instances dedicated to database, web server, and java-based processes. All three of the server instances are backed by Solid State Storage Devices (SSD) and are behind a secure firewall with security groups to manage traffic access to the systems.

- Database instance: optimized for compute power and used to store and manage data for ingest and delivery of data streams that drive web-delivered products on the NERACOOS website, as well as the database that drives the Drupal Content Management System. The server is configured to run PostgreSQL with PostGIS and MySQL database capability.
- Web server server instance: optimized for a general-purpose mix of compute and memory allocation. This instance serves the code that runs the Drupal Content Management System, legacy data-driven applications and products that make up the NERACOOS.org website, and a variety of scripts to manage data ingress/egress. The server is configured for PHP and JavaScript frameworks.
- Java instance: optimized for Tomcat and Java based software packages and is the primary resource for data storage systems such as THREDDS, ERDDAP, Hyrax OPeNDAP, Geoserver, and Docker systems.

The NCDMS 2.0 system leverages the direction of the open-source software industry toward modular, scalable, and containerized architecture. The infrastructure is optimized to define all elements of the infrastructure as code that can be managed and orchestrated with industry standard processes. The NCDMS 2.0 system is housed on an AWS Virtual Private Cloud, and is initially provisioned with Elastic Kubernetes Service (EKS) leveraging two high capacity EC2 instances with shared access to data and file services (AWS Relational Database Service - RDS, AWS Elastic File Storage - EFS) that can be scaled up or down as needed. The development workflow is to create and deploy applications and data integration processes in individual containers on a shared cluster.

The NCDMS 2.0 VPC is firewall protected with security groups to control inbound and outbound traffic access to individual resources. The system is also configured with a constellation of monitoring systems (e.g., CloudWatch for AWS systems, Grafana for visualizing data system analytics for ERDDAP/ArgoCD, Prometheus for measuring and storing data system metrics, Sentry for application and error tracking, UptimeRobot for website health, and Dagster dashboard for data ingress/egress workflows). These tools will reduce risk, minimize downtime and improve efficiency for troubleshooting and fixing issues. Each element of the data management system is managed as a containerized application through an individual Github repository that can be directly deployed to the AWS EKS cluster using automation tools (e.g., Github Actions).

Initial components of the NCDMS 2.0 include the Mariners' Dashboard application and back-end data system (Buoy Barn), ERDDAP v2.18, THREDDS 5.4 with ncWMS, Dagster task orchestration workflow for data ingress/egress to data stores, and an OGC API framework for catalog, query, and retrieval of model data output (pygeoapi, STAC - Spatio-Temporal Asset Catalog, and EDR - Environmental Data Retrieval).

The legacy NERACOOS DMS and NCDMS 2.0 configurations are outlined in the tables below:

Table 1: Legacy NERACOOS DMS hardware and software configuration

Database Instance

AWS Instance type/OS	Detail	Application environment/ software stack	Features
C3.xlarge Ubuntu 14.04 LTS	4 vCPU 7.5 GB memory, 2 x 40 SSD storage (GB)	PostgreSQL with PostGIS and MySQL	Compute optimized: High Frequency Intel Xeon E5-2680 v2 (Ivy Bridge) Processors, enhanced networking, SSD-backed instance storage

Webserver Instance

AWS Instance type/OS	Detail	Application environment/ software stack	Features
C3.large Ubuntu 14.04 LTS	2 cpu, 3.75 GB memory, 2 x 16 SSD storage (GB)	PHP, JavaScript	Compute optimized: High Frequency Intel Xeon E5-2680 v2 (Ivy Bridge) Processors, enhanced networking, SSD-backed instance storage

Java Instance

AWS Instance type/OS	Detail	Application environment/ software stack	Features
M3.large Ubuntu 14.04 LTS	2 vCPU, 7.5 GB memory, 1 x 32 SSD storage (GB)	Tomcat/Java for ODPDX, THREDDS, ERDDAP, GeoServer	General purpose/optimized: High Frequency Intel Xeon E5-2670 v2 (Ivy Bridge) Processors, SSD-backed instance storage

Table 2: NERACOOS DMS 2.0 Configuration

AWS Instance type/OS	Detail	Application environment	Features

Two: M5.xlarge Linux/Unix	2 vCPU, 16 GB memory, EBS SSB storage, EBS bandwidth to 4,750 Mbsp	Kubernetes cluster running Docker containers for web applications, data storage, and data ingress/egress processes	High Frequency Intel Xeon Platinum 8000 Series processor, Turbo CPU clock speed of 3.1 GHz

AWS Service	Features
Elastic Kubernetes Service (EKS)	Cluster node and orchestration of Kubernetes
AWS Relational Database Service (RDS)	PostgreSQL Database for Mariners' Dashboard and Buoy Barn
Elastic File System (EFS)	Flat file storage for ERDDAP/THREDDS
Elastic Container Service (ECS)	EC2 Container management service

Monitoring Service	Description
AWS CloudWatch	Monitors health and usage of AWS systems and services
Grafana	Visualization of data from Prometheus
Prometheus	Monitors and stores data metrics for ERDDAP, ArgoCD
Sentry	Captures and tracks errors encountered by users in web applications (e.g., remote data source not found, server errors)
UptimeRobot	Monitors uptime of website and individual web pages
Dagster Dashboard	Data ingress/egress task orchestration and data workflow monitor

6 NERACOOS Data Streams

Overview

NERACOOS funds a significant amount of data collection in the Northeast region. While we strive for consistency and efficiency, each funded data provider has slightly different internal procedures and systems for data acquisition, processing, QA/QC and managing data. In this section we provide detail on each NERACOOS funded data provider who currently contributes data to the system. The table below (Table 2) provides and overview of the funded data provider, supported assets, and current and future data flow protocols. More detail for each data provider is described in detail in this section.

NERACOOS is working with all funded data providers to produce CF 1.6 netCDF, which will enable the IOOS compliant ncSOS services directly from the NERACOOS TDS. For some providers, particularly UMaine, this has been a fairly complex upgrade due to the differences in discrete sampling geometry conventions for multiple depth and time variables at one station.

Table 3: NERACOOS data provider's data flow overview.

Data Provider	Sensors / Assets	Present Data Flow	Planned Data Flow
University of Maine	7 Buoys in Gulf of Maine	UMaine produces data in netCDF 1.4 and transmits to Web Accessible Folder (WAF), NERACOOS retrieves via rsync and inserts into OBS DB, converts to 1.6 and updates NERACOOS THREDDS catalog and adds data to ERDDAP. NERACOOS sends data to NDBC via FTP.	UMaine will produce netCDF 1.6 for RT and historic data, .nc files will continue to be transmitted to Web Accessible Folder. NERACOOS will retrieve data from WAF and add to NERACOOS THREDDS and ERDDAP. UMaine has also installed an ERDDAP server on site which NERACOOS may point to for updates and direct data access. NERACOOS will make UMaine data available to NDBC via ERDDAP protocol
	3 HF Radar Stations	UMaine sends data to HF RADAR DAC, NERACOOS acquires data via DAC feed and application	No change
	4 Slocum G2 Gliders	UMaine sends data to Rutgers RUCool processing center, Rutgers provides real time QC and forwards data to IOOS Glider Data Assembly Center	UMaine performs real time QC on site, and sends data directly to IOOS Glider DAC
	Satellite	UMaine acquires satellite data from NOAA and NASA data centers, processes data and provides satellite imagery to NERACOOS.	No change
University of Connecticut (UConn)	3 Buoys in LIS (Western LIS, Execution Rocks, and Central LIS)	UConn makes data available through on-site ERDDAP. NERACOOS acquires data from remote UCONN ERDDAP for inclusion in NERACOOS ERDDAP. UConn sends data to NDBC via FTP service	NERACOOS will make UCONN data available to NDBC via ERDDAP protocol
University of New Hampshire (UNH)	3 Buoys/ 1 Shore station (Great Bay Buoy, CML station, CO2 buoy)	Great Bay Buoy, CML, CO2: UNH produces netCDF 1.6 and makes available via on-site THREDDS. NERACOOS acquires data from the UNH THREDDS and adds to ERDDAP. NERACOOS sends data to NDBC via FTP.	NERACOOS will make UNH data available to NDBC via ERDDAP protocol
	1 CDIP	CDIP buoy telemeters data to CDIP Data Center; NERACOOS	No change

		acquires RT data from RSS feed produced by UCSD.	
University of Rhode Island (URI)	Stations of the Narragansett Bay Fixed Site Monitoring Network	Legacy SOS installed on local data system at URI, NERACOOS acquires RT data from seven stations and inserts into OBS DB	URI has installed on-site THREDDS server and netCDF files from these buoys are available. NERACOOS will retrieve data from the URI THREDDS and integrate directly into a ERDDAP. NERACOOS will make URI data available to NDBC via ERDDAP protocol
Woods Hole Oceanographic Institution	4 HF Radar stations	WHOI sends data to HF RADAR DAC, NERACOOS acquires data via DAC feed and application	No change
Charybdis Group	3 Tide Gauges (Scituate, Gloucester, Hampton)	Data are telemetered to sensor provider data management (HOBOLINK) system. NERACOOS acquires data from web service and returns data as .csv for integration into ERDDAP.	No change
Woods Hole Group	2 CDIP Buoys (Cape Cod Bay and Buzzards Bay) 1 HADCP Cape Cod Canal	CDIP buoys telemeters data to CDIP Data Center; NERACOOS acquires RT data from RSS feed produced by UCSD. HADCP telemeters data to NOAA CO-OPS;NERACOOS acquires data from CO-OPS data service	No change

6.1 University of Maine Moored Buoy Array

Summary

The University of Maine (UMaine) operates seven oceanographic data buoys for NERACOOS: A01, B01, E01, F01, I01, M01, N01. The standard measurements on these buoys include near-surface winds, air temperature, barometric pressure, visibility, non-directional waves, surface ocean temperature and salinity, ocean temperature and salinity at depth and ocean currents throughout the water column. Dissolved oxygen and nitrate sensors are deployed on select individual buoys. Additional detail on the parameters measured and specific sensors deployed is available in the NERACOOS Funded Assets table (https://neracoos1.org/wp-content/uploads/2022/11/NERACOOS-Funded-Assets-2.pdf). The data from all instruments are received by a Campbell Scientific CR1000X data logger in each buoy for collection and transmission via a cellular modem or an Iridium satphone modem in the buoy. Backup transmissions are also received hourly via NOAA's GOES satellite systems. All real-time data are processed and quality-controlled by UMaine and provided to NERACOOS. Data are stored in netCDF files; flat ascii files are generated from the netCDF files. These buoys are at least serviced annually and some buoys are serviced twice a year.

Data Flow

The data processing system for data generated by the NERACOOS oceanographic buoys operated by UMaine consists of both field components and shore side components. A programmable data logger inside the buoy well timestamps data received from sensors, aggregates it, and transmits the data to UMaine via two methods, a) cellphone or satellite modem, or b) NOAA GOES satellite. This results in two streams of largely redundant data, designated “realtime” and “goes-realtime” respectively.

The two streams of buoy telemetry received at UMaine are processed by a near-real time processing system that has been in operation since approximately 2001, and extensively modified since then. The system comprises a UNIX-based application server and a UNIX-based web server, and is driven by a number of unix shell scripts, python scripts, and MATLAB scripts. Data is stored and archived in netCDF format, following CF and COARDS conventions in effect when the system was built. Additional metadata for netCDF files and directory structure is accessed via MySQL databases.

NERACOOS receives the provisional “realtime” buoy data from UMaine through an automated process that aggregates data and generates netCDF files that are pushed to an FTP site. On the NERACOOS side, a script runs every 10 minutes that checks for new data packages on the FTP server and acquires files via rsync. NetCDF files go through two processes on the NERACOOS side: legacy and new. The legacy processes use scripts to unpack and extract data and metadata and insert into the Observations Database (OBS DB), which supports a set of operational data products. The newer processes extract metadata from the netCDF files, converts names to CF 1.6 accepted standard, and produces new netCDF CF 1.6 files, which are added to the NERACOOS THREDDS and ERDDAP catalogs. The netCDF files are aggregated by sensor packages and deployment. The reason for redundancy is to maintain functionality of legacy/operational data products as we transition the data system to the netCDF/THREDDS/ERDDAP model. Once all data products are getting data from the ERDDAP services, the legacy OBS DB and associated processes will not be needed. UMaine is also working to produce CF 1.8 compliant files, which will eliminate the conversion step by GMRI. The netCDF files UMaine is developing will also be NCEI 2.0 compliant, which is the required format for archiving. We anticipate that the system for generating CF 1.8 files will be completed in December 2023. The CF 1.8 compliant files will be generated in parallel with the existing legacy netCDF files. The CF 1.8 compliant files merge redundant data streams (cellphone, satellite, post-recovery, as described above), into a single "best available" data stream available to NERACOOS and accessible via rsync or local ERDDAP server.

UMaine also generates historic data files after a buoy is recovered and data are downloaded from the data loggers. Data recovered from data loggers are reconciled with the real-time data files at UMaine and new files are generated that represent a more complete record of that deployment with a higher level of QA/QC. These netCDF files are pushed out to the FTP site into a subdirectory for historic data. The NERACOOS rsync process checks these directories as well, and any new packages are acquired and added to the THREDDS/ERDDAP catalog using the same format described above. In the catalog, they are identified as historic to differentiate from the real-time provisional files.

Quality Control

All data variables reported to NERACOOS in real time netCDF files have an ancillary QC flag. A non-zero flag represents an invalid data value for that record. Malformed data and invalid data structures from instrumentation (similar to “Syntax Test” in QARTOD manuals) are likely to be rejected at the data logger onboard the buoy by embedded software. Shore side, initial quality control checks tailored to the separate data streams (similar to “Timing/Gap” and “Syntax” checks in QARTOD manuals) are performed on all of the cellphone and GOES satellite telemetry. After this point in the processing, the same processing routines are used for both streams.

A range check (“Gross Range Test” per QARTOD manual) is carried out for all reported data variables processed through the system. The “valid_range” netcdf variable attribute is used for this purpose at an early stage in processing. QC flags for data values falling outside the “valid_range” attribute are set to the value “1”. Valid ranges for data variables are set by template prior to buoy testing and deployment. The ranges are determined by buoy personnel based on instrument limitations, empirical data, and knowledge of site environments -- most are tighter than the maximum “sensor span” for a given data variable.

A “Location Check” is also performed for each reported data variable in the system – as GPS positions are updated, the distance from the updated position to the nominal deployment position is calculated and if this is larger than a watch circle radius, alarms immediately notify data management personnel. Data from a confirmed position outside the watch circle are not reported and a QC flag is set to a non-zero value.

In addition to the real-time automated QC checks, data is reviewed by data management personnel on a daily basis using a variety of diagnostic tools including comparison with neighboring platforms (similar to QARTOD “Neighbor Tests”). If data from a sensor is found to be erratic or suspect in any way, the QC flags for associated variables are set to the non-zero value of “2” until the data can be assessed manually for validity. Additional automated and operator-supervised QC is carried out when sensors with onboard storage are recovered post-deployment.

Collaborating scientists provide fluorometers (chlorophyll and turbidity sensors) deployed on UMaine buoys. The collaborating scientists maintain the sensors according to manufacturer specifications and have sensors serviced by the manufacturer. The scientist will deliver and mount the sensors on the buoy and provide UMaine with calibration coefficients prior to deployment. UMaine runs basic automated QC (range and location checks) on the real-time data streams. The collaborating scientists also monitor the data and inform UMaine if they detect any issues.

Collaborating scientists also provide nitrate sensors deployed on one UMaine buoy. The collaborating scientists maintain the sensors according to manufacturer specifications and have sensors serviced and calibrated by the manufacturer. The scientist will deliver the sensors to the UMaine buoy group, who mount sensors on the buoy and provide power and communications cables prior to deployment. UMaine runs basic automated QC (range and location checks) on the real-time data streams. The collaborating scientists also monitor the data and inform UMaine if they detect any issues.

Automated QC checks for the in-situ current measurements (RDI ADCPs) are limited and are targeted for improvement. Part of this limitation is due to restricted bandwidth for telemetry that does not allow for sending back extensive diagnostic data in real-time (i.e. echo intensities, correlation magnitudes, and vertical and error velocities) for acoustic samples.

Automated QC checks for in-situ surface wave measurements (Summit accelerometer, UMaine Directional Wave Module) are also targeted for improvement.

The system under development at UMaine that produces CF 1.8 compliant files, as described in the Data Flow section above, provides primary and secondary QC flags using the IOC/UNESCO/QARTOD flagging standard as recommended by QARTOD project manuals. These flags will be generated in addition to the existing legacy QC flags, and QC information from existing historical files will be mapped into the recommended QARTOD format for primary and secondary QC flags.

Status of required QC tests for the data variables listed in the QARTOD manuals are listed in the following table. Data that fail a test are flagged with a non-zero QC flag.

In-situ temperature and Salinity		**Temperature**	**Salinity**	**Conductivity**
Test 1	Timing/Gap test	*Yes*	*Yes*	*Yes*
Test 2	Syntax Test	*Yes*	*Yes*	*Yes*
Test 3	Location Test	*Yes*	*Yes*	*Yes*
Test 4	Gross Range	*Yes*	*Yes*	*Yes*
Test 5	Climatology Test	*No*	*No*	*No*
Test 7	Rate of change test	*Yes*	*Yes*	*No*

Wind Data		**Wind Speed**	**Wind Direction**	**Wind Gust**
Test 1	Timing/Gap test	*Yes*	*Yes*	*Yes*
Test 2	Syntax Test	*Yes*	*Yes*	*Yes*
Test 3	Location Test	*Yes*	*Yes*	*Yes*

Test 4	Gross Range	*Yes*	*Yes*	*Yes*
Test 5	Climatology Test	*No*	*No*	*No*

In-situ Surface Wave Data		**Significant Wave Height**	**Dominant Wave Period**	**Avg Wave Direction**
Test 16	Stuck Sensor	*No*	*No*	*No*
Test 17	Operational Frequency Range LF	*No*	*Yes*	*No*
Test 18	Energy	*No*	*No*	*No*
Test 19	Bulk Wave Params	*Yes*	*Yes*	*Yes*
Test 20	Rate of Change	*No*	*No*	*No*

In-situ Current Observations		**Current Speed**	**Current Direction**	**Current U**	**Current V**
Test 2	Check Sum	*No*	*No*	*No*	*No*
Test 3	Sensor Tilt	*No*	*No*	*No*	*No*
Test 4	Speed of Sound	*No*	*No*	*No*	*No*
Test 6	Signal Strength	*No*	*No*	*No*	*No*
Test 10	Current Speed	*Yes*	*Yes*	*Yes*	*Yes*
Test 11	Current Direction	*Yes*	*Yes*	*Yes*	*Yes*
Test 12	Horizontal Velocity	*No*	*No*	*No*	*No*
Test 17	Flat Line	*No*	*No*	*No*	*No*
Test 18	Echo Intensity	*No*	*No*	*No*	*No*

Nutrient Data		**Nutrient concentration**
Test 1	Timing/Gap test	*Yes*
Test 2	Syntax Test	*Yes*
Test 3	Location Test	*Yes*
Test 4	Gross Range	*Yes*
Test 5	Climatology Test	*No*

In-situ Dissolved Oxygen		**Dissolved Oxygen Concentration**	**Dissolved Oxygen Percent Saturation**
Test 1	Timing/Gap test	*Yes*	*Yes*

Test 2	Syntax Test	*Yes*	*Yes*
Test 3	Location Test	*Yes*	*Yes*
Test 4	Gross Range	*Yes*	*Yes*
Test 5	Climatology Test	*No*	*No*

Backup and Archival

The UMaine data management system includes several data backup procedures. As buoy telemetry is received and processed in real-time, netCDF files specific to each buoy deployment are updated in a directory tree on the UMaine application server. Immediately after update of buoy data for a particular buoy, that data is replicated (via rsync) to the web server for distribution.

That data directory on that web server is used by NERACOOS (and CariCOOS) to ingest (via rsync) buoy data.

Once daily the entire directory tree on the application server (which contains historical and post-recovery as well as real-time netCDF data files) is replicated (again, in case any post real-time edits or updates were missed) to the web server for distribution.

Both the application server and the web server are backed up daily by intermittent full backups to rotating hard drives on a backup server located in a different on-campus building at UMaine.

As part of the ongoing upgrade to the UMaine processing system, described in the Data Flow section above, all processing and distribution of UMaine observing system data will be moved to virtual machines which can be replicated or operated offsite or in a cloud environment. Virtual machines will be periodically backed up via replication to insure a spare operational system is available on short notice. Hourly and daily data directory replication as carried out currently will continue.

Data archival with NCEI is described in Section 8.3.

6.2 University of Maine High Frequency Radar

Summary

UMaine operates three High Frequency Radar (HFR) stations for NERACOOS. These stations are located on Grand Manaan Island, New Brunswick, Cape St. Mary, Nova Scotia, and Green Island, Maine.

Data Flow

SeaSonde SSRS-100 hardware and software from CODAR Ocean Sensors run autonomously at the three sites. The remote sites acquire HFR data, archive the raw data, generate radial files, and push the radial files to a processing server at UMaine.

Raw data is collected and brought back to UMaine during visits to the remote sites. The processing server, an SSDP-100 Central Site Management/Data Combining Station from CODAR Ocean Sensors, applies QA/QC and archives the radials files received from remote stations. The HFR radial files are ingested from the UMaine processing server by HRFNET, the national distributed processing system, where additional QA/QC is carried out and spatially averaged surface current data ("totals files") are generated. HFRNET makes radial and total files available through THREDDS and provides data visualization products for the aggregated HFR data.

Quality Control

UMaine relies on manufacturer recommended software settings in the SeaSonde software acquisition software running on remote HFR sites, for acquisition of HFR data and unattended generation and quality control of the radials files prior to ingest by HFRNET. Further filtering and QA/QC of radials files, as well as generation and QA/QC of spatially averaged surface current data is carried out by HFRNET.

Backup and Archival

HFR radials files from UMaine are archived by the HFRNet at UCSD, along with totals files generated by HRFNET from the radial files. UMaine maintains a single set of the RAW HFR data collected from site visits on individual hard drives at the Orono campus.

6.3 University of Maine Satellite Imagery

Summary

The Satellite Oceanography Data Laboratory (SODL) at UMaine provides the NERACOOS Satellite data products. UMaine collects, processes, QA/QCs, and archives the satellite data and delivers high-resolution data products to NERACOOS. Coverage includes the Gulf of Maine and Long Island Sound. The primary sources of satellite data for the NERACOOS stream are: NOAA AVHRR and MODIS (Aqua and Terra) which SODL uses to provide fully processed sea surface temperature (AVHRR) and ocean color and sea surface temperature (MODIS) products multiple times each day at 1 km resolution.

Data Flow

UMaine SODL acquires and processes multiple satellite data streams to produce NERACOOS satellite data products. Two sources of satellite data are collected by the UMaine system for the NERACOOS data products:

1. MODIS (Moderate Resolution Imaging Spectroradiometer) is an instrument aboard NASA's EOS Aqua and Terra satellites. MODIS Aqua and Terra views the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands (channels) that are used to calculate ocean surface temperature and biogeochemical products (SODL produces and serves a surface chlorophyll a product).
2. AVHRR (Advanced Very High Resolution Radiometer) aboard the NOAA POES satellite system is a 5-channel radiometer providing measurements used to calculate sea surface temperature. Currently, NOAA maintains two such

operational satellites, providing at a minimum 4 overpasses per day (depending on orbital geometry). Each satellite covers the Earth surface every 1-2 days.

Each orbit of NOAA AVHRR SST data and NASA MODIS color/SST data that include coverage of the Gulf of Maine are received, archived, and processed into geophysical products (4-6/day AVHRR, 1-2/day MODIS) at 1km resolution.

The SODL is linked via wide bandwidth Internet connection to the main US east coast NOAA satellite data server facility to receive the raw, unprocessed, NOAA HRPT telemetry stream that contains the AVHRR data in near real time (http://www.class.ngdc.noaa.gov/saa/products/welcome). The HRPT data acquired by the SODL originate from the Wallops Island VA reception facility.
For the MODIS data stream, the SODL server system interrogates the NASA Goddard Space Flight Center MODIS satellite reception and archive system where SODL has a subscription (https://earthdata.nasa.gov/earth-observation-data/near-real-time/download-nrt-data/modis-nrt).

Data Processing and QA/QC
MODIS: MODIS raw swath data are processed to fully cloud-masked and atmospherically corrected bio-geophysical ocean products using SODL-developed scripts that call subroutines from NASA's SEADAS satellite data processing software (https://seadas.gsfc.nasa.gov). Ancillary data files and updated coefficients required for the processing are acquired in real time from additional NASA data servers within the SEADAS processing system. Both the chlorophyll a and the SST MODIS products produced for NERACOOS utilize the standard global NASA ocean product algorithms implemented within SEADAS and are updated as NASA updates their products.

AVHRR
AVHRR raw data are processed to SST products using SeaSpace's proprietary TERASCAN software (https://www.seaspace.com). For this processing, each scene is manually viewed, checked and adjusted by a trained SODL operator to ensure the image geometry is correct. Two products are created, 1) a "Standard cloud-masked" product that uses information from within the scene itself and coefficients developed by SODL and implemented within the SeaSpace TERASCAN software, and 2) an "Enhanced cloud masked" product that uses the Standard Cloud mask product plus additional information from SST data over the previous several days implemented within SODL-written scripts modified from protocols published by Cayula and Cornillon, 1995 (Journal of Atmospheric and Oceanic Technology, 12: 821-829).

Backup and Archival
AVHRR SST final image products are archived in TERASCAN data format (TDF, a proprietary format used by the SeaSPACE TERASCAN processing software). These files,

however, are accessible / readable as simple binary files. Additionally, JPEG copies of each processed SST image are produced, posted to the UMaine SODL web site, and archived.

MODIS products are archived as netCDF files.

Both sets of products are archived in a hierarchical directory structure at the UMaine Advanced Computing Group (https://acg.umaine.edu), a professionally monitored, maintained, high performance computational facility with multiple back-up.

6.4 University of Connecticut Moored Buoy Array

Summary

University of Connecticut (UConn) operates three oceanographic data buoys for NERACOOS: Execution Rocks (EXRX), Western Long Island Sound (WLIS), and Central Long Island Sound (CLIS). These moorings measure wind speed and direction, salinity, conductivity, temperature, pressure, currents, and dissolved oxygen near the surface, bottom and mid-depth. The WLIS and CLIS buoys also support wave observations. Additional detail on the parameters measured and specific sensors deployed is available in the NERACOOS Funded Assets table (https://neracoos1.org/wp-content/uploads/2022/11/NERACOOS-Funded-Assets-2.pdf). All buoys are on a yearly service cycle with an approximate turnaround of 3-4 weeks. During this time the buoys are hauled and returned to the Avery Point facility. Mooring hardware is replaced and the buoy hull and tower is repaired and repainted. Instrument sensors are cleaned, calibrated, and repaired if needed, and any software upgrades applied.

Data Flow

The data from each sensor is collected by a Campbell Scientific CR1000 datalogger for storage and transmission via cellphone TCP/IP data transfers. All data received at UConn are automatically entered into a POSTGRESQL database and quality-controlled. UConn generates and sends data messages to NDBC as part of the real-time data processing stream. NERACOOS retrieves UConn data hourly from NDBC and incorporates into the NERACOOS OBS DB where they are available to NERACOOS data products.

UConn is currently connecting their ERDDAP server directly to the POSTGRESQL database for data access. ERDDAP provides the ability to convert this data to netCDF format for the end user as well as the ability for remote ERDDAP servers to directly link to these data access links.

Quality Control

The key component of the data collection and dissemination process aboard the metocean buoys is the data logger. The data logger collects, and transmits the data, via cellular communications, back to UConn where it is stored and subsequently accessed and redistributed by NERACOOS. Nearly all the data are sampled and collected every 15 minutes and published to the web in near real time.

All data are initially received as comma delimited ASCII data files that are then read by automated scripts and stored in a POSTGRESQL database as provisional data. First order QA/QC protocols are subsequently applied to the data from the provisional database. UConn is using ERDDAP to convert data to netCDF format for end user distribution.

Quality control checks are applied to the data after insertion into the. The main function of the first check is to find invalid data ranges from the sensors. Invalid data ranges are removed and replaced with a NAN. The data are then scanned for duplicate timestamps and location verification. Duplicate data are removed from the database, data collected outside the defined station (watch) radius are saved but stored in a separate data table. Data are then manually reviewed and compared with any additional data collected from or near the station (i.e., CTD casts, other sampling surveys, etc.) and if determined to be suspect, is flagged.

Backup and Archival
The live data stream collected using the proprietary data logger software and stored as comma delimited ASCII files are backed up to a local external hard drive and onto UConn's Enterprise File Services network drive, located offsite on the Storrs campus, every hour. In addition to the live data stream, data is logged on each instrument and on the data logger. These data are collected when the buoy is hauled or the instruments are retrieved and subject to the same backup procedure. Data on the database server is backed up onto two separate external hard drives once daily.

Data archival with NCEI is described in Section 8.3.

6.5 University of New Hampshire Mooring and Sensor Network

Summary
The University of New Hampshire (UNH) operates several monitoring assets for NERACOOS including the Great Bay buoy, CO2 buoy, the Coastal Marine Lab system (CML) and a CDIP buoy. The Great Bay Buoy is a highly instrumented mooring in Great Bay, NH that is deployed during ice-free months. This buoy measures a standard suite of meteorological and oceanographic parameters including air temperature, winds, water temperature and salinity. The Great Bay buoy also monitors dissolved oxygen, chlorophyll, optical properties, nitrates, pH, and CO2. CML is located at the mouth of the Great Bay Estuary. This station monitors surface meteorology and water quality and CO2. The CO2 buoy is located off the coast of NH and measures surface meteorology, water temperature, salinity, and air and water CO2. The CDIP buoy is located offshore on Jeffrey's Ledge and monitors waves and water temperature. Additional detail on the specific sensors deployed is available in the NERACOOS Funded Assets table (https://neracoos1.org/wp-content/uploads/2022/11/NERACOOS-Funded-Assets-2.pdf).

Data Flow

Great Bay Buoy

The data from all instruments on the Great Bay buoy are connected to Campbell Scientific data loggers. Data are telemetered via cellular modem to a shore-based computer at the UNH Durham campus. Real-time data from each sensor are automatically processed hourly to align data in time and produce hourly means, medians, minimums, maximums, and standard deviations. Data also undergo automated quality control tests (QARTOD tests) when available. After quality control testing, data are transmitted to NERACOOS as text files, and are also provided via THREDDS server in netCDF format.

CML

Data from the CML are continuously logged by a computer in the CML building, and transmitted hourly to a computer system on the UNH Durham campus, where data from each sensor are automatically processed hourly to align data in time and produce hourly means, medians, minimums, maximums, and standard deviations. Data also undergo automated quality control tests (QARTOD tests) when available. After quality control testing, data are transmitted to NERACOOS as text files, and are also provided via THREDDS server in netCDF format.

CO2 Buoy

The data from all instruments are connected to an Axys Watchman500 data logger with the exception of the xCO2 and pH readings, which log directly to the MAPCO2 system and are sent back to PMEL daily. The remaining data are telemetered via radio transmission to a shore-based computer at the Seacoast Science Center and transmitted hourly to a computer system on the UNH Durham campus. Data undergoes automated quality control tests (QARTOD tests) when available. After quality control testing, data are transmitted to NERACOOS as text files, and are also provided via THREDDS server in netCDF format.

UNH maintains a THREDDS server and is producing CF 1.6 netCDF files for the CML, CO2, and the Great Bay buoy data streams. NERACOOS acquires data directly from THREDDS to insert into the NERACOOS ERDDAP. The NERACOOS THREDDS server is configured to point to UNH's THREDDS server to populate the NERACOOS data catalog.

The UNH maintained CDIP buoy data flow is described in Section 6.9.

Quality Control

All CML, CO2 and Great Bay Buoy data variables reported to NERACOOS as netCDF files, that have published QARTOD procedures, include ancillary QC flags corresponding to the appropriate QARTOD procedure. Real time text file data have values flagged with a value other than 1 (i.e. "Suspect" or "Fail") removed. A table of checks performed on appropriate variables is given below.

A “Location Check” is also performed for each reported data variable in the CO2 Buoy system – as GPS positions are updated, the distance from the updated position to the nominal deployment position is calculated and if this is larger than a watch circle radius, alarms immediately notify data management personnel. Data from a confirmed position outside the watch circle are not reported and a QC flag is set to a non-zero value.

The CO2 buoy CO2 sensor data is sent to Pacific Marine Environmental Lab (PMEL) to undergo QA/QC. The QA/QC procedure used at PMEL is described in http://www.earth-syst-sci-data.net/6/353/2014/essd-6-353-2014.html

All chlorophyll sensors operated by UNH are returned to the manufacturer (WetLabs) and recalibrated on the recommended schedule. UNH staff visits the CO2 buoy four times per year to collect discrete chlorophyll samples that are used to validate the chlorophyll sensor operation. The Great Bay NERRS system wide monitoring program collects extensive water quality data from Great Bay monthly and these results are used to validate the chlorophyll sensor on the Great Bay buoy.

In addition to the real-time automated QARTOD QC checks, data are reviewed by data management personnel on a regular basis using a variety of diagnostic tools including comparison with neighboring platforms. Additional automated and operator-supervised QC is carried out when sensors with onboard storage are recovered post deployment. UNH runs QARTOD tests on the CO2 buoy, CML, and Great Bay buoy chlorophyll data.

Status of required QC tests for the data variables listed in the QARTOD manuals are listed in the following table. Data that fail a test are flagged with a non-zero QC flag.

CO2 Buoy

Temperature/salinity			
		Temperature	Salinity
Test 1	Gap Test	*yes*	*yes*
Test 2	Syntax Test	*yes*	*yes*
Test 3	Gross Range Test	*yes*	*yes*
Test 4	Climatology Test	*yes*	*yes*
Oxygen		Oxygen Concentration	
Test 1	Gap Test	*yes*	
Test 2	Syntax Test	*yes*	
Test 3	Gross Range Test	*yes*	
Test 4	Climatology Test	*yes*	
Turbidity		Turbidity	

Test 1	Gap Test	*yes*			
Test 2	Syntax Test	*yes*			
Test 3	Gross Range Test	*yes*			
Test 4	Climatology Test	*yes*			

Atmospheric data		Wind Speed	Wind Direction	Wind Gust	Air Temperature
Test 1	Gap Test	*yes*	*yes*	*yes*	*yes*
Test 2	Syntax Test	*yes*	*yes*	*yes*	*yes*
Test 3	Gross Range Test	*yes*	*yes*	*yes*	*yes*
Test 4	Climatology Test	*yes*	*yes*	*yes*	*yes*

CML Station

Temperature/salinity

		Temperature	*Salinity*
Test 1	Gap Test	*yes*	*yes*
Test 2	Syntax Test	*yes*	*yes*
Test 3	Gross Range Test	*yes*	*yes*
Test 4	Climatology Test	*yes*	*yes*

Oxygen		*Oxygen Concentration*
Test 1	Gap Test	*yes*
Test 2	Syntax Test	*yes*
Test 3	Gross Range Test	*yes*
Test 4	Climatology Test	*yes*

Turbidity		*Turbidity*
Test 1	Gap Test	*yes*
Test 2	Syntax Test	*yes*
Test 3	Gross Range Test	*yes*
Test 4	Climatology Test	*yes*

Atmospheric data		*Wind Speed*	*Wind Direction*	*Wind Gust*	*Air Temperature*
Test 1	Gap Test	*yes*	*yes*	*yes*	*yes*
Test 2	Syntax Test	*yes*	*yes*	*yes*	*yes*
Test 3	Gross Range Test	*yes*	*yes*	*yes*	*yes*
Test 4	Climatology Test	*yes*	*yes*	*yes*	*yes*

Great Bay Buoy

Temperature/salinity

		Temperature	Salinity
Test 1	Gap Test	*yes*	*yes*
Test 2	Syntax Test	*yes*	*yes*
Test 3	Gross Range Test	*yes*	*yes*
Test 4	Climatology Test	*yes*	*yes*

Oxygen		Oxygen Concentration
Test 1	Gap Test	*yes*
Test 2	Syntax Test	*yes*
Test 3	Gross Range Test	*yes*
Test 4	Climatology Test	*yes*

Turbidity		Turbidity
Test 1	Gap Test	*yes*
Test 2	Syntax Test	*yes*
Test 3	Gross Range Test	*yes*
Test 4	Climatology Test	*yes*

Atmospheric data		Wind Speed	Wind Direction	Wind Gust	Air Temperature
Test 1	Gap Test	*yes*	*yes*	*yes*	*yes*
Test 2	Syntax Test	*yes*	*yes*	*yes*	*yes*
Test 3	Gross Range Test	*yes*	*yes*	*yes*	*yes*
Test 4	Climatology Test	*yes*	*yes*	*yes*	*yes*

Chlorophyll		
		Chlorophyll
Test 1	Gap Test	*yes*
Test 2	Syntax Test	*yes*
Test 3	Gross Range Test	*yes*
Test 4	Climatology Test	*yes*

Oxygen		*Oxygen % saturation*
Test 1	Gap Test	*yes*
Test 2	Syntax Test	*yes*
Test 3	Gross Range Test	*yes*
Test 4	Climatology Test	*yes*

xCO2		*xCO2*
Test 1	Gap Test	*yes*
Test 2	Syntax Test	*yes*
Test 3	Gross Range Test	*yes*
Test 4	Climatology Test	*no*

Other water data		*pH*	*Nitrate*	*PAR*	*fDOM*
Test 1	Gap Test	*yes*	*yes*	*yes*	*yes*
Test 2	Syntax Test	*yes*	*yes*	*yes*	*yes*
Test 3	Gross Range Test	*yes*	*yes*	*yes*	*yes*
Test 4	Climatology Test	*no*	*no*	*no*	*no*

Backup and Archival
All UNH data are archived on the local UNH computer, an external hard drive, and backed up to a cloud data storage service (Box@UNH).

Data archival with NCEI is described in Section 8.3.

6.6 University of Rhode Island

Summary
NERACOOS provides partial support to the University of Rhode Island (URI) to operate two dock and five seasonal mooring buoy-based water quality sensors and implement data management processes for these stations that are part of the Narragansett Bay Fixed Site Monitoring Network (NBFSMN). NERACOOS also funds nutrient monitoring at the dock stations and provides the real-time data access for all seven stations. The partners of the NBFSMN fund most of the data collection and management effort. Data from this network are being integrated into the NERACOOS DMS. Additional detail on the specific sensors deployed is available in the NERACOOS Funded Assets table (https://neracoos1.org/wp-content/uploads/2022/11/NERACOOS-Funded-Assets-2.pdf).

Data Flow
At the present time, real-time data are acquired from seven stations in the NBFSMN via an SOS service. Data are acquired and inserted into the OBS DB every hour where they are available to data products. As part of the transition to the new data management system, URI has set up a THREDDS server where they are making netCDF files directly available for those two year-round dock stations and five seasonal stations. The NERACOOS TDS server is configured to point to URI's TDS and populate the catalog. The NBFSMN buoy data are available through the NERACOOS ERDDAP server. The list of stations currently serving surface and bottom data via OPeNDAP and SOS are Mount View, Quonset Point, Conimicut Pt, North Prudence, Poppasquash Point, Greenwich Bay (year-round and the GSO-Pier (year-round).

Quality Control
The NBFSMN has an extensive Quality Assurance Project Plan (QAPP) that documents their QA procedures and was approved by the EPA. Please see NBFSMN QAPP (https://neracoos1.org/wp-content/uploads/2022/11/nbfsmn_QAPP_2020.pdf).

Backup and Archival
Data from the NBFSMN are backed up on the base station computer at Marine Ecosystems Research Laboratory building within URI-GSO. All NBFSMN data (raw, edited and corrected data sets) are stored and available to the public annually through the RIDEM website (http://www.dem.ri.gov/programs/emergencyresponse/bart/stations.php).

6.7 Charybdis Group Tide Gauges

Summary

NERACOOS supports the operation of three tide gauges that are located in Scituate, MA, Gloucester, MA, and Hampton, NH. These gauges monitor coastal water level. The Charybdis Group out of Boston, Massachusetts, maintains the stations.

Data Flow

A programmable Microwave Radar transducer (MRT) is coupled with an integrated data logger/telemetry unit (DLT) to upload the data over the 3G cellular network to a cloud based hosted service and archive (HAS).

NERACOOS currently uses an ftp system to retrieve data in .csv format from a ftp server. The data are currently accessed on demand when requested by the data products. The data access system is moving to a SOAP based SOS service, which will greatly improve the process for acquiring and disseminating the data.

Quality Control

The MRT has an inherit precision and accuracy that is specified by the manufacturer. In accordance with the manufacturer the instrument is sent back to the factory for evaluation and calibration at the prescribed intervals. The manufacturer of the MRT is ultimately responsible for the accuracy of the instrument as long as the maintenance intervals are respected.

Waves and swells affect the measurement of MLLW elevation; such effect is removed by applying an appropriate sampling protocol (NOAA or CG30) as described above. This process also takes care of the occasional presence of flotsam.

The location and placement of the tide gauge is carefully studied to ensure the least interference of local currents and other dynamic phenomena. The Platform Of Opportunity (POO) is selected to be a stable platform and its elevation is surveyed at the time of installation and yearly thereafter. If a change in elevation of the platform is discovered and no specific point in time can be determined for when the shift occurred (such as an extreme episodic event), then the change in elevation is considered gradual and is linearly applied retroactively to the previous year's data; alternatively, the change is applied as a single vertical shift from the date of the identified episodic event, and the data are shifted accordingly from that point onward. Charybdis Group LLC is responsible for the elevation reference of the tide gauge.

For any transmission error from the DLT to the HAS we defer to the Cellular Network and host CRC (Circular Redundancy Checks). At the specified 6 minute acquisition rate, the DLT is capable of holding up to 2 years worth of data. In the event of a cellular network disservice the DLT will be able to locally store the data and later upload to the HAS asynchronously. The data integrity from the DLT to the HAS is the responsibility of the DLT Manufacturer, since the HAS is a bundled service.

The MRT measures raw distance to water (ullage). As well, the DLT stores raw distance to water. The HAS stores raw distance to water and converted water level with respect to MLLW; the conversion formula is hosted by the HAS.

Backup and Archival
All data is maintained on the HAS on raid 6 and higher disk arrays and periodically archived in accordance to best industry standards.

6.8 Woods Hole Oceanographic Institution High Frequency Radar

Summary
The Woods Hole Oceanographic Institution (WHOI) operates four HFR stations within the southwestern component of the Gulf of Maine on NERACOOS's behalf. Three of the stations are located within the Massachusetts Bay and provide surface current coverage of the bulk of the Bay and especially the areas around Stellwagen Bank with spatial resolutions of 1-2 km. A fourth, long-range station is located north of Massachusetts Bay close to the Massachusetts/New Hampshire border.

Data Flow
The data from all stations is transmitted to WHOI where it undergoes QA/QC processing. The output of the WHOI data processing method is an industry standard ascii text file that follows the existing and required data file formats and metadata requirements of NOAA-IOOS and is transferred to the HFRNet at UCSD for ingestion into the total velocity calculation that is managed and distributed by HFRNet.

Quality Control
The WHOI provided HF radar based observations of surface currents are processed from the raw spectral estimates of radar backscatter into surface currents following similar methods to those used and documented by Codar Ocean Sensors for the SeaSonde HF radar with a few key differences. Following Kirincich et al. (2012)[1] and de Poalo et al. (2015)[2], the raw results of the direction finding step are subjected to advanced quality control thresholds to eliminate suspect data and processed to spatially averaged surface currents using power-weighted radial averages. These two steps, in addition to the longtime basis for the spectral estimates themselves have been shown to lead to higher quality radial velocity results.

Backup and Archival
HFR data generated by WHOI are archived at the HFRNet at UCSD.

[1] Kirincich, A. R., T. de Paolo, and E. Terrill (2012), Improving HF radar estimates of surface currents using signal quality metrics, with application to the MVCO high-resolution radar system, J. Atmos. Ocean. Technol., 29(9), 1377–1390, doi: 10.1175/JTECH-D-11-00160.1.

[2] de Paolo, T. D., E. Terrill, and A. Kirincich, (2015): Improving SeaSonde Radial Velocity Accuracy and Variance using Radial Metrics. Trans. Ocean. MTS/IEEE Genova, 6.

6.9 Woods Hole Oceanographic Institution Glider

Summary

The Woods Hole Oceanographic Institution (WHOI) operates a Slocum glider in the waters of the northern and western Gulf of Maine during the winter to survey for North Atlantic right whales and other baleen whales while simultaneously collecting data from a conductivity-temperature-depth instrument and an optical fluorometer/turbidity sensor. The glider uses the WHOI developed digital acoustic monitoring (DMON) instrument to detect, characterize and classify low-frequency tonal sounds of whales, which are transmitted to shore and reviewed by an analyst. All data, including the analyst's estimates of whale occurrence, are made available on a publicly accessible website, robots4whales.whoi.edu.

Data Flow

Sound detection, temperature, conductivity, fluorescence and turbidity data are collected while the glider is submerged. Every 2 hours, the glider comes to the surface and transmits these data, as well as glider engineering data, to a computer on shore at WHOI via the Iridium satellite communication system. These data are immediately transferred to a server where they are processed and displayed graphically on the robots4whales.whoi.edu website. An analyst reviews the acoustic detection data twice a day, and the results of that analysis (i.e., estimates of whale occurrence) are also posted on the robots4whales.whoi.edu site. A simple application programming interface (API) makes the whale occurrence estimates available to other applications, including, for example, Whale Alert, Whale Map, Whale Safe, and Mysticetus, and whale detection information is also shared directly with stakeholders via email and text messages.

Quality Control

Detection information transmitted to shore is displayed graphically on the robots4whales.whoi.edu website. Modulation in frequency and time of each sound detection is depicted as a pitch track (analogous to notes on a page of sheet music), which is interpreted in a similar way to an audio spectrogram. The analyst takes into account the shape and amplitude of the pitch track, the classification information, and how isolated the pitch track is from other pitch tracks to determine whether the sound was produced by a whale or not. Pitch track information is reviewed in nominal 15-minute periods, and whale occurrence estimates are based on the number and (for some species) pattern of pitch tracks. The review process and protocol is documented at robots4whales.whoi.edu/#protocol, and is described in Baumgartner et al., 2019, Persistent near real-time passive acoustic monitoring for baleen whales from a moored buoy: system description and evaluation, Methods in Ecology and Evolution 00:1-14.

Backup and Archival
All of the acoustic and environmental data are archived at both WHOI and the NOAA Northeast Fisheries Science Center, and estimates of whale occurrence are shared publicly on both the robots4whales.whoi.edu site and the NOAA NEFSC Passive Acoustic Cetacean Map (https://apps-nefsc.fisheries.noaa.gov/pacm). NOAA NEFSC is participating in the development of the Passive Acoustic National Data Assembly Center, and all data collected to date will be automatically archived in this DAC when it goes live.

6.10 CDIP Wave Buoys

Summary
There are four CDIP buoys deployed in the NERACOOS region: Jeffrey's Ledge in the Gulf of Maine (maintained by UNH), Cape Cod Bay and Buzzards Bay (maintained by the Woods Hole Group) and Block Island in Rhode Island (maintained by USACE). The buoys record wave height, wave period, wave direction and water temperature. NERACOOS partially supports the UNH maintained buoy and fully supports the Cape Cod and Buzzards Bay buoys. NERACOOS ingests data from the CDIP program for all four buoys.

Data Flow
The data for all four CDIP buoys are managed and disseminated by the central data processing and archiving facility at University of California San Diego. The buoy telemeters data via satellite to the CDIP data center. There are failover processes as well that use HF radio to shore stations. Data are retrieved from the data logger after recovery and transmitted to the CDIP Data Center. NERACOOS access the real-time data directly from the CDIP ERDDAP (http://erddap.cdip.ucsd.edu/erddap) server and makes it available through data products on the website.

Quality Control
QC for the CDIP buoy data are conducted by the CDIP program.

Backup and Archival
The CDIP program provides data backup and archival.

7 Data Sharing and Access

Data served through the NERACOOS website (http://www.neracoos.org) complies with the NOAA Data Sharing Procedural Directive by providing open access to real-time and historic data through a variety of services including data products and machine-readable services. The data and metadata formats used by NERACOOS adhere to IOOS guidance based on open standards (CF 1.6, ncSOS, ISO 19115). The website provides access to several data services that support machine-to-machine data access as well as a variety of data products that allow end users to view and/or download near real-time data, historic data, and model predictions in a variety of formats. The NERACOOS website also provides access to information about NERACOOS including but not limited to strategic framework, board of directors, members, by-laws, grants, news, projects and partners.

The NERACOOS website utilizes a Responsive Web Design approach that enables data products and information to be effectively viewed on mobile devices.

Figure 4: Screen capture of the NERACOOS home page.

7.1 Mariners' Dashboard

The NERACOOS Mariners" Dashboard Data product (https://mariners.neracoos.org/) provides access to real-time ocean observations from regional observing assets including NERACOOS buoys and sensors, NDBC buoys, C-MAN stations, CDIP buoys, Environment Canada buoys, CO-OPS stations, NERRS stations and various short term research buoys in the region. This product provides a map of the region with clickable symbols that identify the location of ocean observing stations. Selecting a station provides the user a view of the latest weather and sea surface observations from that station. The product also includes access to selected forecast output associated with each station location. NERACOOS is in the process of implementing a model viewer capability within the Mariners Dashboard. The model viewer allows a user to select a model output field and generate an interactive map view of the output field.

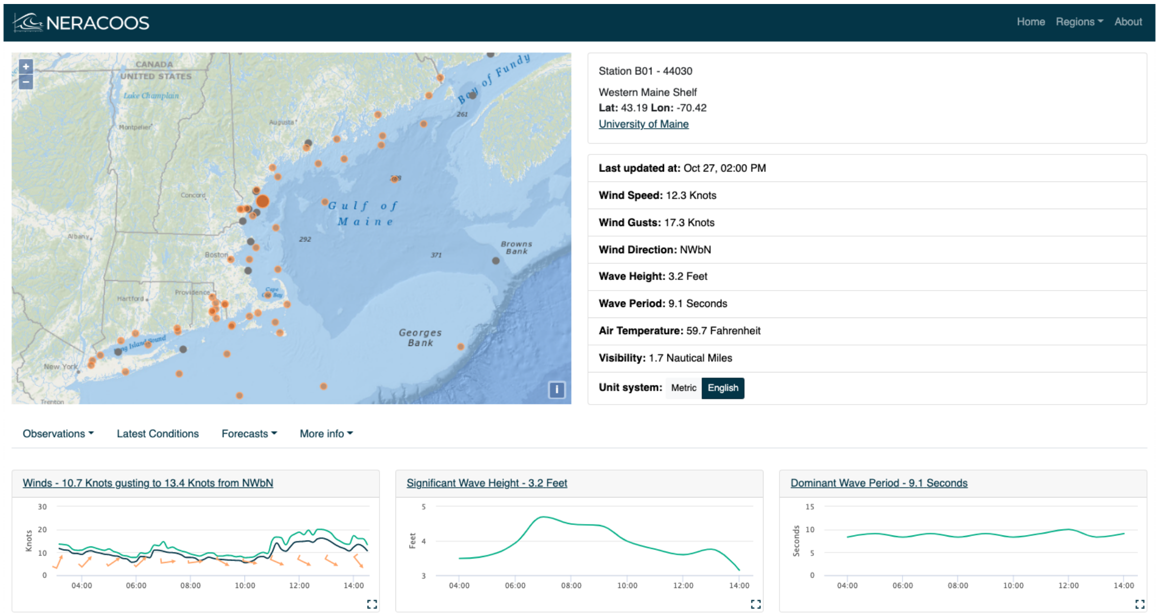


Figure 5: Screen capture of the NERACOOS Mariners' Dashboard.

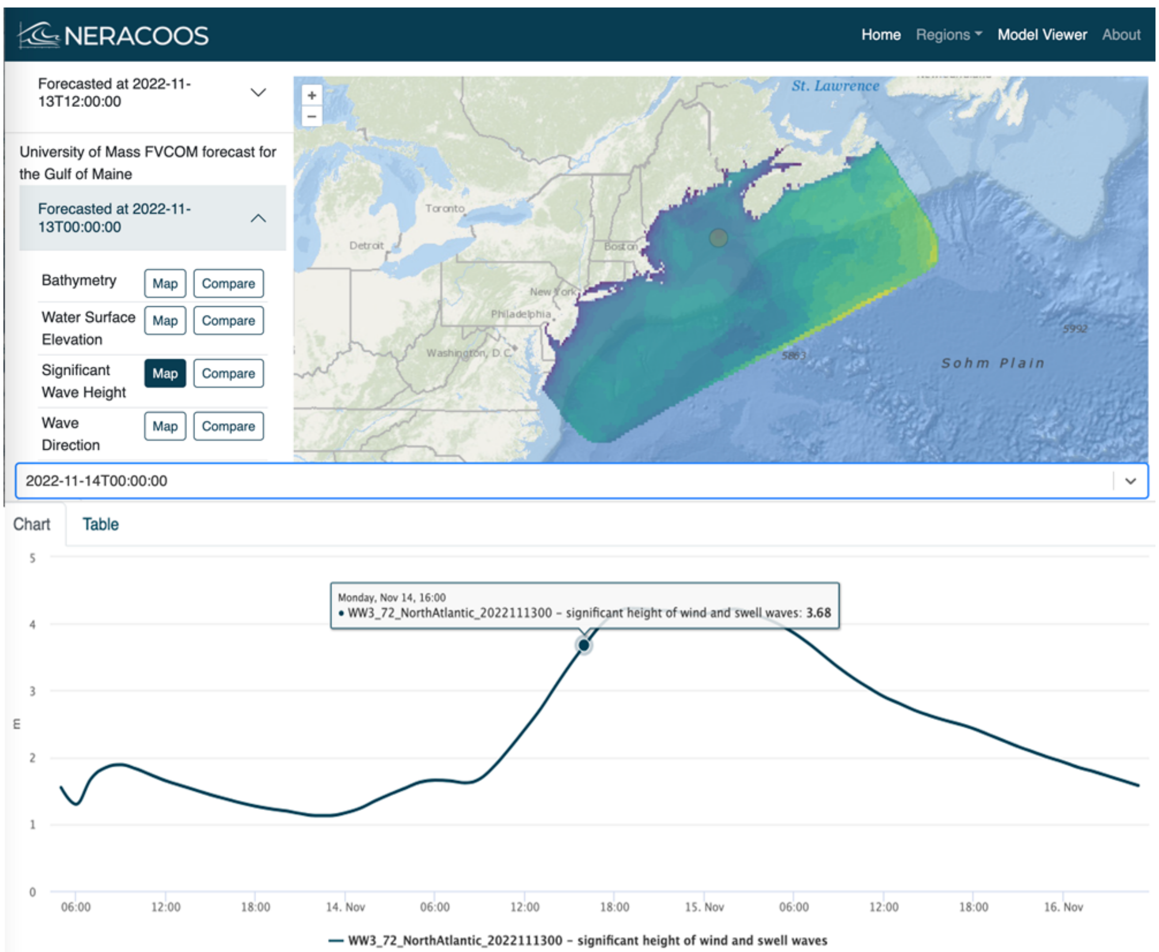


Figure 6: Screen capture of model viewer product

7.2 Ocean Climate

The Ocean Climate product allows a user to view the latest buoy observations in comparison to a 20 year climatology for the same meteorological or oceanographic parameter. This allows the user to see how current conditions compare to the average conditions for that location. The tool allows a user to look from year to year and view how any given years data compares to the average conditions.

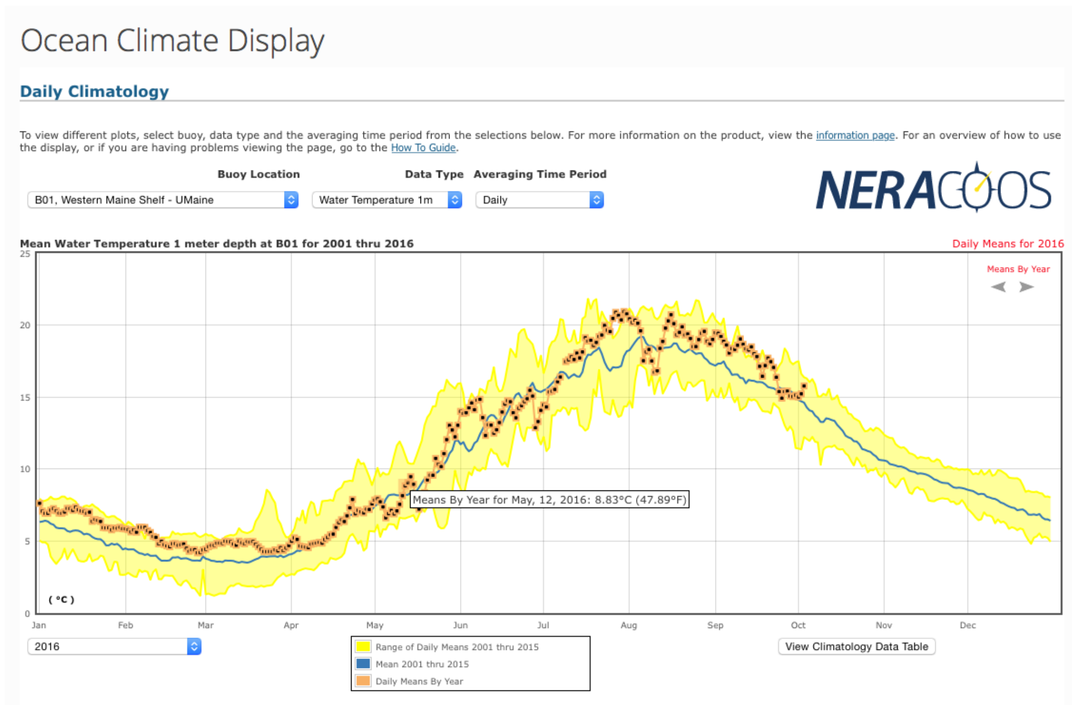


Figure 7: Screen capture of the NERACOOS ocean climate product.

7.3 Graphing and Download

The graphing and download tool provides a user-friendly interface to graph and/or download hourly meteorological and oceanographic data from buoys and stations in the NERACOOS region. Graphing and Download allows a user to select the parameter(s) they are interested in, select a time frame they want to see/access the data for, select the buoy location(s) they want data from and then they can view a graph of the data or access the data in several common file formats.

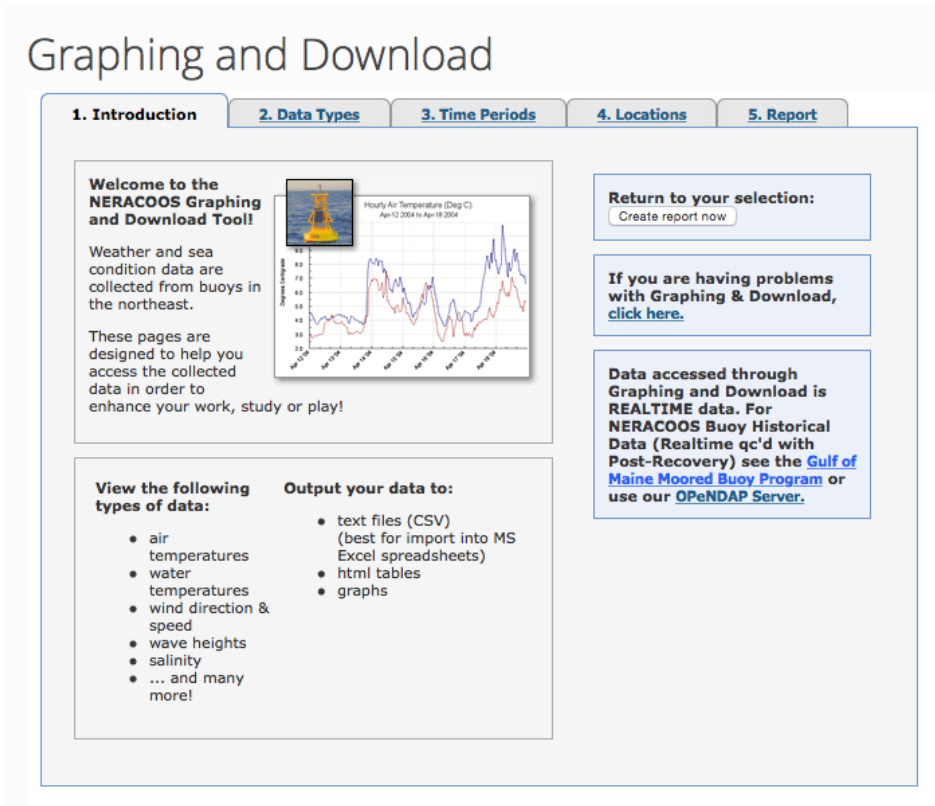


Figure 8: Screen capture of the NERACOOS graphing and download product.

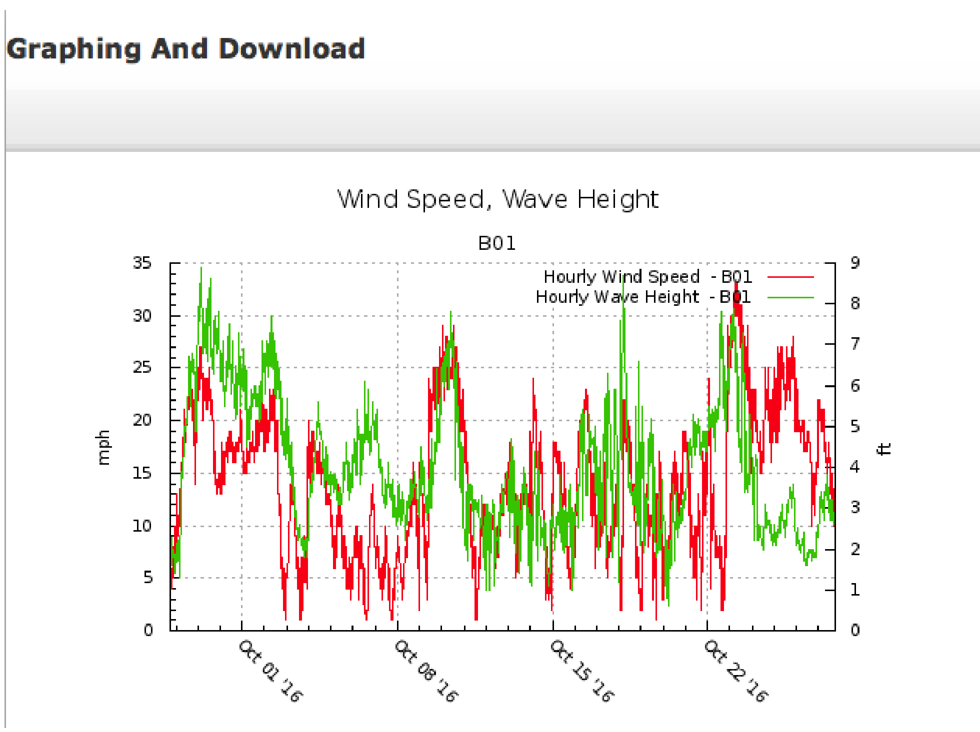


Figure 9: Screen capture of the NERACOOS graphing and download output.

7.4 Satellite Imagery

The Satellite Oceanography Data Laboratory at UMaine provides a product that allows users to view satellite imagery of sea surface temperature and chlorophyll concentration.

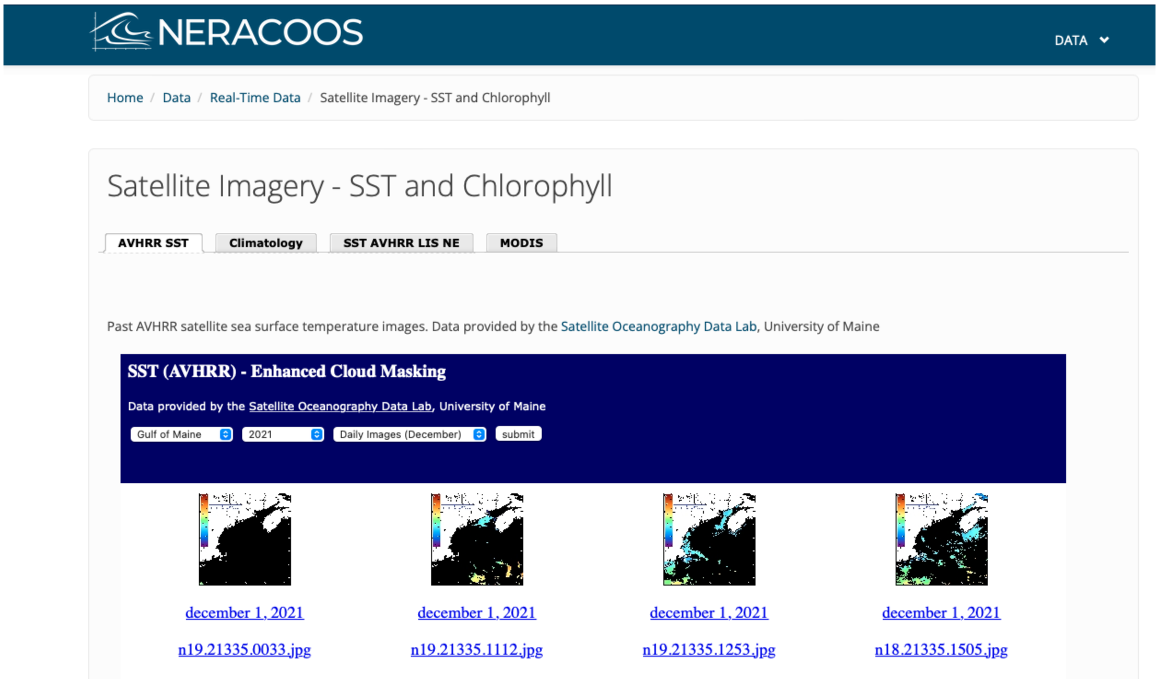
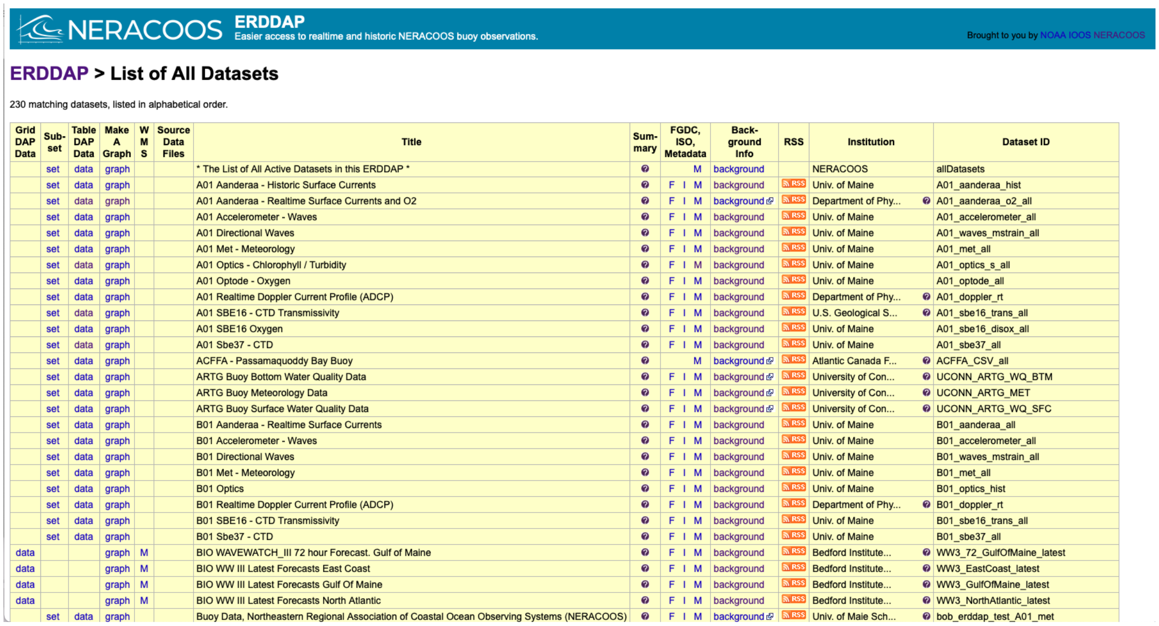


Figure 10. Screen capture of the NERACOOS satellite imagery product.

7.5 NERACOOS ERDDAP

NERACOOS maintains an ERDDAP (the Environmental Research Division's Data Access Program) data server (http://www.neracoos.org/erddap/index.html) that provides a simple, consistent way to download subsets of scientific datasets in common file formats and make graphs and maps. ERDDAP also serves as the backend for some of our data products such as the Mariners' Dashboard



NERACOOS ERDDAP
Easier access to realtime and historic NERACOOS buoy observations.
Brought to you by NOAA IOOS NERACOOS

ERDDAP > List of All Datasets

230 matching datasets, listed in alphabetical order.

Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	W M S	Source Data Files	Title	Sum-mary	FGDC, ISO, Metadata	Back-ground Info	RSS	Institution	Dataset ID
	set	data	graph			* The List of All Active Datasets in this ERDDAP *	?	M	background		NERACOOS	allDatasets
	set	data	graph			A01 Aanderaa - Historic Surface Currents	?	F I M	background	RSS	Univ. of Maine	A01_aanderaa_hist
	set	data	graph			A01 Aanderaa - Realtime Surface Currents and O2	?	F I M	background	RSS	Department of Phy... ?	A01_aanderaa_o2_all
	set	data	graph			A01 Accelerometer - Waves	?	F I M	background	RSS	Univ. of Maine	A01_accelerometer_all
	set	data	graph			A01 Directional Waves	?	F I M	background	RSS	Univ. of Maine	A01_waves_mstrain_all
	set	data	graph			A01 Met - Meteorology	?	F I M	background	RSS	Univ. of Maine	A01_met_all
	set	data	graph			A01 Optics - Chlorophyll / Turbidity	?	F I M	background	RSS	Univ. of Maine	A01_optics_s_all
	set	data	graph			A01 Optode - Oxygen	?	F I M	background	RSS	Univ. of Maine	A01_optode_all
	set	data	graph			A01 Realtime Doppler Current Profile (ADCP)	?	F I M	background	RSS	Department of Phy... ?	A01_doppler_rt
	set	data	graph			A01 SBE16 - CTD Transmissivity	?	F I M	background	RSS	U.S. Geological S... ?	A01_sbe16_trans_all
	set	data	graph			A01 SBE16 Oxygen	?	F I M	background	RSS	Univ. of Maine	A01_sbe16_disox_all
	set	data	graph			A01 Sbe37 - CTD	?	F I M	background	RSS	Univ. of Maine	A01_sbe37_all
	set	data	graph			ACFFA - Passamaquoddy Bay Buoy	?	M	background	RSS	Atlantic Canada F... ?	ACFFA_CSV_all
	set	data	graph			ARTG Buoy Bottom Water Quality Data	?	F I M	background	RSS	University of Con... ?	UCONN_ARTG_WQ_BTM
	set	data	graph			ARTG Buoy Meteorology Data	?	F I M	background	RSS	University of Con... ?	UCONN_ARTG_MET
	set	data	graph			ARTG Buoy Surface Water Quality Data	?	F I M	background	RSS	University of Con... ?	UCONN_ARTG_WQ_SFC
	set	data	graph			B01 Aanderaa - Realtime Surface Currents	?	F I M	background	RSS	Univ. of Maine	B01_aanderaa_all
	set	data	graph			B01 Accelerometer - Waves	?	F I M	background	RSS	Univ. of Maine	B01_accelerometer_all
	set	data	graph			B01 Directional Waves	?	F I M	background	RSS	Univ. of Maine	B01_waves_mstrain_all
	set	data	graph			B01 Met - Meteorology	?	F I M	background	RSS	Univ. of Maine	B01_met_all
	set	data	graph			B01 Optics	?	F I M	background	RSS	Univ. of Maine	B01_optics_hist
	set	data	graph			B01 Realtime Doppler Current Profile (ADCP)	?	F I M	background	RSS	Department of Phy... ?	B01_doppler_rt
	set	data	graph			B01 SBE16 - CTD Transmissivity	?	F I M	background	RSS	Univ. of Maine	B01_sbe16_trans_all
	set	data	graph			B01 Sbe37 - CTD	?	F I M	background	RSS	Univ. of Maine	B01_sbe37_all
data			graph	M		BIO WAVEWATCH_III 72 hour Forecast. Gulf of Maine	?	F I M	background	RSS	Bedford Institute... ?	WW3_72_GulfOfMaine_latest
data			graph	M		BIO WW III Latest Forecasts East Coast	?	F I M	background	RSS	Bedford Institute... ?	WW3_EastCoast_latest
data			graph	M		BIO WW III Latest Forecasts Gulf Of Maine	?	F I M	background	RSS	Bedford Institute... ?	WW3_GulfOfMaine_latest
data			graph	M		BIO WW III Latest Forecasts North Atlantic	?	F I M	background	RSS	Bedford Institute... ?	WW3_NorthAtlantic_latest
	set	data	graph			Buoy Data, Northeastern Regional Association of Coastal Ocean Observing Systems (NERACOOS)	?	F I M	background	RSS	Univ. of Maie Sch... ?	bob_erddap_test_A01_met

Figure 11: Screen capture of the NERACOOS ERDDAP list of data sets.

7.6 Website Analytics

NERACOOS uses Google analytics to monitor Activity on our website. Google analytics allows us to track overall usage and look at the usage of specific products as well as how users navigate within our website.

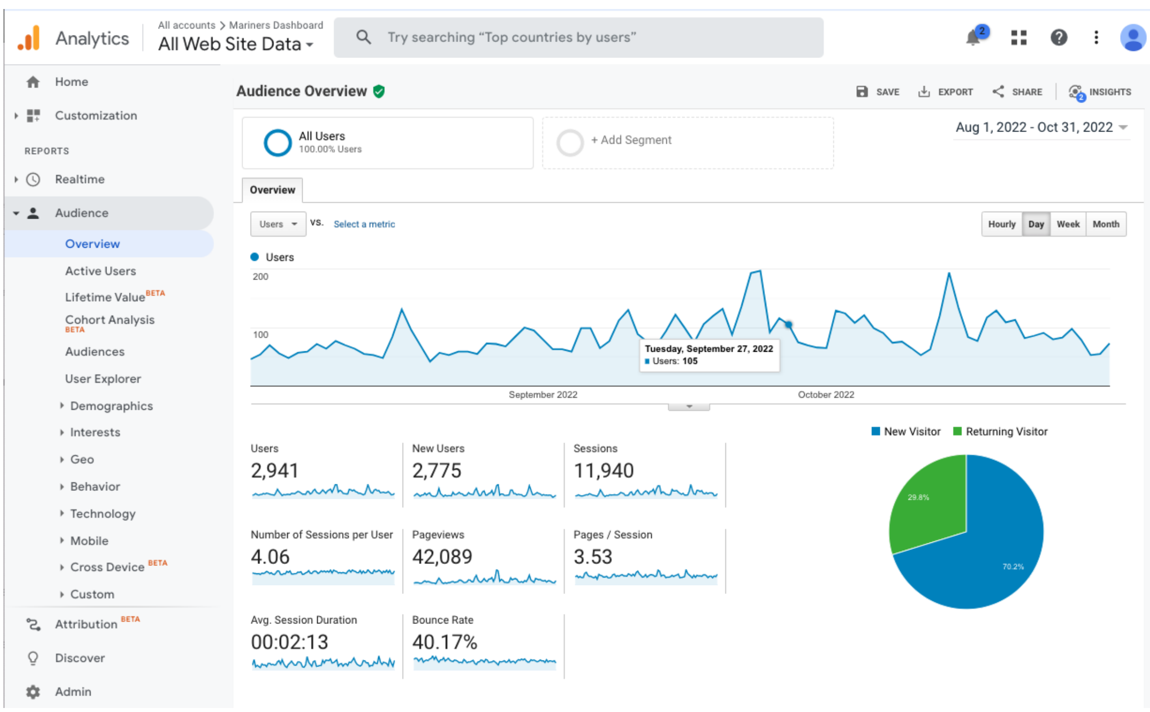


Figure 12: Screen capture of NERACOOS Google Analytics dashboard.

8 Data Backup and Archival

Security of NERACOOS data is ensured using the “here, near, there” paradigm outlined by the Long Term Ecological Research (LTER) guidance on digital research data and documents (http://intranet2.lternet.edu/content/protecting-your-digital-research-data-and-documents). NERACOOS funded data providers keep a local version of the raw data (here); submit data to the NERACOOS data management system (near), and will ultimately archive data with the NCEI (there). The GMRI Data Manager is responsible for working with funded partners to implement processes around data handling, curation and submission to the NERACOOS data management system and NCEI.

8.1 Data Provider Backup

NERACOOS funded data providers maintain local versions of NERACOOS data and backup data using their organization's backup systems, which are described in section 6. The table below provides an overview of the data providers local backup systems.

Table 4: Data provider data backup procedures.

Data Provider	Backup Protocol
Charybdis Group	All data are maintained on a cloud based hosted service and archive on raid 6 and higher disk arrays and periodically archived in accordance to best industry standards.
University of Rhode Island (URI)	URI data files are backed up on the base station computer at Marine Ecosystems Research Laboratory building within URI-GSO.
University of Connecticut (UConn)	Data files are stored on a Raid 6 data array as well as backed up to a local external hard drive at UConn Avery Point Campus and onto UConn’s Enterprise File Services network drive, located offsite on the UConn’s Storrs campus.
University of Maine	Once daily the entire directory tree on the application server (which contains historical and post-recovery as well as real-time netCDF data files is replicated to the web server for distribution. Both the application server and the web server are backed up daily by incremental backup and weekly by full backup to rotating hard drives on a backup server located in a different on-campus building.
University of New Hampshire (UNH)	Data files are archived on the local UNH computer, an external hard drive, and backed up to a cloud data storage service (Box@UNH).

8.2 NERACOOS Data System Backup

The data stored on the NERACOOS Amazon EC2 environment is backed up using Amazon Simple Storage Service (S3). This includes netCDF files from providers, data stored in the observation databases and shared file systems, application code, and website content. The scalable, secure solution provides nightly incremental backups of the entire Amazon EC2 instances. GMRI manages the account on Amazon EC2 and has full access to configure and deploy instances and manage the back up systems.

Data are retained over a 30-day revolving window. From this scalable, secure solution data can be recovered from a point in time during that window and restored quickly in the event of a system failure.

The collection of software code that includes data processing scripts, web services and web-delivered applications and products is managed internally through version control systems. GMRI uses Git, which is managed, in an online private repository on GitHub (https://github.com/). GMRI maintains a library of technical documentation of current systems on the GitHub repository as well as legacy documentation on Google Docs. GMRI also maintains a public NERACOOS GitHub repository (https://github.com/neracoos-open). These documents include detailed descriptions of the Amazon Web Services configuration, protocols for disaster recovery, data framework configuration and product development.

8.3 Archival at NCEI

NERACOOS is currently working with NCEI to establish a process of submitting data to NCEI. NERACOOS data providers will create netCDF files compatible with IOOS CF 1.6 and NCEI netCDF Templates v2.0 data formats and metadata standards. Historical data will be produced in yearly or monthly files and made available on an http accessible web site by the data providers. NERACOOS has established a Web Accessible Folder (WAF) to allow NCEI to pull the data as needed. Each data provider's data will be submitted to NCEI via the single NERACOOS WAF. Newer, near real-time files will be created monthly and stored on the WAF site. NCEI will pull these files monthly. NERACOOS will provide guidance and tools to the data providers to ensure their netCDF files meet the standards required.

We are implementing this system in a phased approach and starting with UMaine. The NCEI Request to Archive has been initiated with the University of Maine, which oversees the buoy program for both NERACOOS and CariCOOS. Please see the 2017 NCEI Request to Archive document (https://drive.google.com/file/d/1rrkefqDO14lCCfQyMFYyjQ7kXiaWyLLu/view?usp=sharing) and the 2017 NERACOOS Submission Agreement Document (https://drive.google.com/file/d/1bGJGmuaUDToshR6bMehA1xHHRIB7rl82/view?usp=sharing). We are currently working with NCEI to update these documents.

At this time, the University of Maine effort has successfully archived historical data for one of the seven buoys. The data template that is CF and IOOS metadata compliant will be applied to the other 6 buoys for full archiving. This effort is complementary to the work to transform datasets for QARTOD and NDBC compliance so it will be completed in early 2023. The other data partners are in the process of transforming data for compliance to CF, IOOS Metadata, and QARTOD standards. Upon completion, we will leverage the outcome of the UMaine work with NCEI to provide a model for our other data providers. We anticipate the system to submit data from UMaine to NCEI will be fully implemented by February 2023. After UMaine is completed we will work with UNH to implement their archival process and then work with UConn to implement their process. We anticipate that the UNH and UConn archival implementation will be complete by June 2023.

9 Appendix I: NERACOOS Data Source Table

Provider Type Key: NF = NERACOOS Funded, F/S = Federal/State, P = Private

Observations

Provider	**Type**	**Locations**	**Provider Type**		
			NF	**F/S**	**P**
UMaine Buoys	Real Time Historic	A01 Mass Bay B01 Western Maine Shelf E01 Central Maine Shelf F01 Penobscot Bay I01 Eastern Maine Shelf M01 Jordan Basin N01 Northeast Channel *C01 Casco Bay *J01 Eastport *K01 Saint Johns, NB *L01 Yarmouth, NS (*buoys no longer deployed, historical records available)	X		
UMaine HF Radar	Real Time	Grand Manaan Island, New Brunswick Cape St. Mary, Nova Scotia Green Island, Maine	X		
UConn Buoys	Real Time Historic	44022 Execution Rocks Long Island Sound 44039 Central Long Island Sound 44040 Western Long Island Sound 44060 Eastern Long Island Sound LDLC3 New London Ledge Light	X		X

URI Buoys and Sensors	Real Time Historic	Narragansett Bay - Mount View Narragansett Bay - Quonset Point			X
UNH Buoys and Sensors	Real Time Historic	CML Coastal Marine Lab Field Station CO2 Appledore Island GREAT_BAY Great Bay, NH	X		
Charybdis Group Tide Gauges	Real Time Historic	Hampton, NH Tide Gauge Gloucester, MA Tide Gauge Scituate Tide, MA Gauge	X		
WHOI HF Radars	Real Time	Race Point, MA Eastern Point, MA Fourth Cliff, MA Salisbury, MA	X		
WHOI Sensors	Real Time Near Real Time	MVCO SeaNode Martha's Vineyard, MA ESP HAB - several locations, Gulf of Maine			X
NOAA NEFSC Sensors	Historic	eMOLT - various locations, Gulf of Maine		X	
NOAA NDBC Buoys and Stations	Real Time Historic	44005 Cashes Ledge 44007 Casco Bay 44008 Nantucket 44011 Georges Bank 44013 Boston Harbor 44017 23 Nautical Miles SW of Montauk Point, NY 44018 SE Cape Cod 44020 NANTUCKET SOUND 44027 Jonesport, ME CMAN Stations BUZM3 Buzzards Bay * IOSN3 Isle of Shoals * MDRM1 Mt Desert Rock * MISM1 Matinicus Rock *		X	
NERRS Sensors	Real Time Historic	BGXN3 Great Bay Reserve, NH NAXR1 Naragansett Bay Reserve, RI NAXR1 Narragansett Bay Reserve, RI WAXM3 Waquoit Bay Reserve, MA WEXM1 Wells Reserve, ME		X	
Environment Canada Buoys	Real Time	44137 East Scotia Slope * 44150 La Have Bank * 44258 Halifax Harbor *		X	
NOAA CMAN	Real Time	CMAN MDRM1 - Mt Desert Rock CMAN MISM1 - Matinicus Rock CMAN IOSN3 - Isle of Shoals CMAN BUZM3 - Buzzards Bay		X	

CDIP Buoys	Real Time	CDIP 154 Block Island, RI CDIP 160 Jeffreys Ledge (UNH) CDIP 221 Cape Cod Bay CDIP 260 Buzzards Bay	X	X	
SmartBay Canada	Real Time	SB01 Mouth of Placentia Bay, NL, Canada SB04 Placentia Bay, NL, Canada		X	
NOAA NOS CO-OPS Water Level	Real Time	8410140 Eastport, ME 8411250 Cutler Naval Base, ME 8413320 Bar Harbor, ME 8418150 Portland, ME 8419317 Wells, ME 8423898 Fort Point, NH 8443970 Boston, MA 8447386 Fall River, MA 8447387 Borden Flats Light at Fall River, MA 8447930 Woods Hole, MA 8448725 Menemsha Harbor, MA 8449130 Nantucket Island, MA 8447435 Chatham, MA 8452660 Newport, RI 8454049 Quonset Point, RI 8452944 Conimicut Light, RI 8452944 Providence, RI 8461490 New London, CT 8465705 New Haven, CT 8467150 Bridgeport, CT 8510560 Montauk, NY 8516945 Kings Point, NY Scituate Harbor, MA (NWS)		X	
USGS Water Services Water Level	Real Time	Saco, ME Tide Gauge		X	

Models

Provider	Type	Description	NF	F/S	P
UMass	Forecast Historic	NECOFS/FVCOM oceanographic model system Northeast Atlantic and Mass. Bay domains. (via TDS SOS at NERACOOS)	X		
BIO	Forecast	WW3 wave models for Northeast Atlantic Domain. (via TDS SOS at NERACOOS)	X		
UConn	Forecast Nowcast	HFR STPS (nowcast/forecast)	X		
NOAA NOMADS/ NCEP	Forecast	Wind Forecasts for Gulf of Maine (daily via OPenDAP)		X	

NOAA/ NOMADS/ ESTOFS	Forecast	Extratropical Surge and Tide Operational Forecast System		X	

Satellite

Provider	**Type**	**Description**	**NF**	**F/S**	**P**
UMaine	Near Real Time	Sea surface temperature, chlorophyll imagery Historic, climatology	X		X
NOAA Coastwatch	Near Real Time			X	