

NERACOOS Data Management Plan

Version 1.0 – 2/20/2014

[Extracted from Appendix A of the NOAA EDMC Data Management Planning Procedural Directive (2011): https://geo-ide.noaa.gov/wiki/index.php?title=Data_Management_Planning_PD]

This Template is provided as an example to assist DM Plan developers. It may be downloaded from the [DMP Repository](#). The Repository also includes other plans as examples.

Note: Some of this information may be used to populate fields in a formal metadata record about the data collection. These are indicated for illustrative purposes only with [→ field name].

1. General Description of Data to be Managed

- 1.1. Name of the Dataset or data collection project [→ gmd:title].

Northeast Regional Association for Coastal and Ocean Observing Systems

- 1.2. Keywords that could be used to characterize the data, and vocabulary from which those keywords were obtained (e.g., GCMD, CF Conventions, etc.) [→ gmd:MD_Keywords]

aanderaa, buoy, chemistry, chlorophyll, circulation, component, conductivity, current, current_direction data_quality, current_speed data_quality, current_u data_quality, current_v data_quality, currents, data, density, department, depth, direction, direction_of_sea_water_velocity, east, eastward, eastward_sea_water_velocity, height, level, maine, marine, name, neracoos, north, northward, northward_sea_water_velocity, o2, ocean, oceanography, oceans, Oceans > Ocean Chemistry > Chlorophyll, Oceans > Ocean Chemistry > Oxygen, Oceans > Ocean Circulation > Ocean Currents, Oceans > Ocean Optics > Turbidity, Oceans > Ocean Pressure > Sea Level Pressure, Oceans > Ocean Temperature > Water Temperature, Oceans > Ocean Waves > Significant Wave Height, Oceans > Ocean Waves > Wave Period, Oceans > Ocean Winds > Surface Winds, Oceans > Salinity/Density > Conductivity, Oceans > Salinity/Density > Density, Oceans > Salinity/Density > Salinity, optics, oxygen, period, physical, pressure, quality, salinity, school, sciences, sea, sea_water_temperature, sea_water_velocity, seawater, sensor, significant, speed, station, station_name, surface, temperature, temperature data_quality, time, turbidity, university, velocity, water, wave, waves, winds

- 1.3. Summary description of the data to be generated [→ gmd:abstract].

Buoys are collecting meteorological and oceanographic data from the surface to 2500 meters at the deepest location. The data include 10-minute to hourly readings of air temperature, water temperature, salinity, currents, wave height and period,

wind speed and direction. Some buoys are collecting dissolved oxygen, chlorophyll, pH and CO2. Models for the region include FVCOM for the Gulf of Maine, and finer scale estuaries and bays. FVCOM forecast includes water temperature, salinity, waves, current speed and direction. Wave Watch III data for wave height and period is available at several resolutions from the Gulf of Maine to the East Coast.

1.4. Anticipated temporal coverage of the data [→ gmd:EX_TemporalExtent].

Ongoing, some buoys have been reporting since 1999.

1.5. Anticipated geographic coverage of the data [→ gmd:EX_Extent]

The Northeastern Regional Association of Coastal Ocean Observing Systems is a component of the national Integrated Ocean Observing System (IOOS), spanning coastal waters from the Canadian Maritime Provinces to the New York Bight.

1.6. What data types will you be creating or capturing? (e.g., digital numeric data, photographs, video, acoustic records, database tables, spreadsheets, paper records, physical samples, etc.)

Digital numeric data

1.7. How will you capture or create the data? (e.g., satellite, airplane, unmanned aerial system, radar, weather station, moored buoy, research vessel, autonomous underwater vehicle, animal tagging, manual surveys, enforcement activities, etc.)

The data are collected primarily through HF radar, moored buoys.

1.8. Where will this plan be stored electronically besides in the NOAA DMP Repository?

The plan will be stored and displayed on the NERACOOS website and the IOOS website.

1.9. What volume of data is anticipated to be collected in the Project Time Frame?

Currently, the size of the data collected through the NERACOOS system is

1.10. Will the data contain Personally Identifiable Information or any information whose distribution may be restricted by law or national security?

No

2. **Points of Contact** (Give name, title, location, e-mail address, phone number and mailing address, as appropriate.) [→ gmd:CI_ResponsibleParty]

2.1. Who can, or could, represent this data collection project on NOAA's Data Management Integration Team (DMIT)? Current members of DMIT are listed at

https://geo-ide.noaa.gov/wiki/index.php?title=DMIT_Membership.

Ru Morrison of the NERACOOS, represents NERACOOS on the NOAA Data Management Integration Team, ru.morrison@neracoos.org

2.2. Who is the overall point of contact for the data collection?

NERACOOS: Eric Bridger, NERACOOS Data Manager, ebridger@gmri.org

2.3. Who is responsible for verifying the quality of the data?

NERACOOS: Eric Bridger, NERACOOS Data Manager, ebridger@gmri.org

2.4. Who is responsible for answering questions about the data collection?

NERACOOS: Eric Bridger, NERACOOS Data Manager, ebridger@gmri.org

2.5. Who is responsible for data documentation and metadata activities?

NERACOOS: Eric Bridger, NERACOOS Data Manager, ebridger@gmri.org

2.6. Who is responsible for the data storage and data disaster recovery activities?

NERACOOS: Eric Bridger, NERACOOS Data Manager, ebridger@gmri.org

2.7. Who is responsible for ensuring adherence to this data management plan, including ensuring that appropriate resources are available to implement the data management plan?

NERACOOS: Eric Bridger, NERACOOS Data Manager, ebridger@gmri.org

3. Data Stewardship

3.1. What quality control procedures will be employed?

Post-recovery QA/QC is the responsibility of the individual data providers, automated QA/QC protocols for real-time, provisional

3.2. What is the overall lifecycle of the data from collection or acquisition to making it available to customer?

Real-time observations will be uploaded hourly(or other increments?) to the NERACOOS data portal (http://neracoos.org/realtime_map). The final QA/QC'd data will be available on the project partner websites (is this true?) and the NERACOOS website.

4. Data Documentation

- 4.1. Which metadata repository will be used to document this data collection?

Service metadata for all NERACOOS web services and datasets will be stored in the IOOS Data Catalog. Web services are entered into the Data Catalog through the IOOS Service Registration process and the IOOS Data Catalog is currently in development by the IOOS Program Office. All NERACOOS metadata is also available through the Global Change Master Directory (<http://gcmd.nasa.gov/>) and through a web accessible folder linked from the NERACOOS website. Finally, NOAA will include NERACOOS metadata in the NOAA Enterprise Data Inventory (<http://project-open-data.github.io/implementation-guide/>) that is included in the IOOS Data Catalog.

- 4.2. In addition to discovery-level metadata, what additional metadata or other documentation is necessary to fully describe the data and ensure its long-term usefulness? How will that metadata be collected and updated?" Is there a requirement to document this data collection in other metadata repositories?

Metadata are collected and updated by data providers and NERACOOS per operating procedures described in cooperative agreements and other IOOS Program Office guidance. For observational assets supported by NERACOOS, the NERACOOS data manager reviews metadata at the time of initial data feed setup and when the data providers alert the data manager that something changed with the sensors and/or sampling protocol. Missing/incorrect metadata are corrected or the sub-regional data provider is contacted and asked to update their metadata records. The data provider will alert the data manager to these changes. The IOOS Data Catalog will serve as the primary metadata repository for IOOS Data. Metadata: FGDC, ISO19115 for data is exposed via THREDDS ISO, ERDDAP ISO, FGDC.

- 4.3. What standards will be used to represent data and metadata elements in this data collection. *Note:* The [EDMC Data Documentation Procedural Directive](#) calls for the use of ISO 19115 and related standards for data documentation.

Data and metadata elements will be provided in a variety of dialects, including ISO, SOS IOOS templates, SensorML, netCDF with Attribute Conventions Dataset Discovery (ACDD), and THREDDS XML and ERDDAP. All datasets will have their metadata in these standard formats. As NERACOOS web services are registered in the IOOS Service Registry, the IOOS Data Catalog will convert metadata in these multiple formats to ISO 19115.

5. Data Sharing

- 5.1. Will the data be made available to the public? If so, what is the expected date of first availability? Is this a one-time data collection, or an ongoing series of measurements? Will there be a Principal Investigator hold or other delay between data collection and publication, and if so for how long? [Note: the [Data Sharing for NOAA Grants Procedural Directive](#) provides useful guidance for sharing data in a timely manner.]

All NERACOOS Datasets are currently available to the public.

NERACOOS observations are currently available in near real-time on the NERACOOS

website (http://www.neracoos.org/realtime_map) and through the data access protocols listed in Section 5.4.

Historical data from buoy observations are available for download on the NERACOOS website <http://www.neracoos.org/thredds/catalog/UMO/All/Historical/catalog.html>

http://www.neracoos.org/datatools/historical/graphing_download

In addition, these same datasets are archived at NODC on a periodic basis (See Section 7).

- 5.2. If the data are not to be made available to the public, explain why and under what authority distribution may be restricted. [NOAA Administrative Order 212-15](#), "Management of Environmental Data and Information" (2010) states that Environmental data will be visible, accessible and independently understandable to users, except where limited by law, regulation, policy (such as those applicable to personally identifiable information or protected critical infrastructure information or proprietary trade information) or by security requirements.

N/A

- 5.3. Will users be subject to any access conditions or restrictions, such as submission of non-disclosure statements, special authorization, or acceptance of a licensing agreement?

N/A

- 5.4. What data access protocols will be used to enable data sharing? The use of open-standard, interoperable, non-proprietary web services is recommended (for example, OpENDAP, or Open Geospatial Consortium (OGC) web services).

Open-source interoperable web services are the primary means for sharing data. All NERACOOS data is available through IOOS-approved web services, and may be available through other data access protocols. The data access protocols that expose each dataset are listed below.

All historic and near-realtime observation are available via:

THREDDS – OpENDAP, ERDDAP, ncSOS, WMS

Metadata – THREDDS, ncISO, ERDDAP

Various data access portals hosted on the NERACOOS web site:

SOS: http://www.neracoos.org/thredds/UMO_SOS_historical_realtime_agg.html

THREDDS: <http://www.neracoos.org/thredds/catalog.html>

WMS: <http://www.neracoos.org/thredds/wms/WW3/GulfOfMaine.nc?service=WMS&version=1.3.0&request=GetCapabilities>

ERDDAP: <http://www.neracoos.org:8190/erddap/index.html>

- 5.5. In what catalogs will these services or data be made registered to enable discovery by users and other Catalogs?

All NERACOOS web services are registered in the IOOS Service Registry, which will populate an IOOS Data Catalog where services and datasets will be made available for discovery. In addition, IOOS will catalog metadata that is in the IOOS Data Catalog also in the NOAA Enterprise Data Inventory (<http://project-open-data.github.io/implementation-guide/>).

6. Initial Data Storage and Protection

- 6.1. Where and how will the data be stored initially (i.e., prior to being sent to a long-term archive facility)?

Each co-PI will store data for component projects locally at their respective institutions. Data shared with NERACOOS are stored in NetCDF and PostgreSQL DB on NERACOOS AWS Cloud servers:

- **Reserved Amazon Elastic Compute Cloud Instance type c1.xlarge.**
- **8 cpu equivalent processing units with 8 gigabytes of memory.**
- **One 64 gigabyte volume and One 400 gigabyte volume.**
- **Servers run in a virtual private cloud behind a custom security group.**
- **Data is backed up nightly to Amazon S3 with incremental backups going back 3 months.**
- **Code base is in a hosted git repository.**

- 6.2. How will the data be protected from accidental or malicious modification or deletion? Discuss data back-up, disaster recovery/contingency planning, and off-site storage relevant to the data collection.

Each co-PI will have a mechanism for storing a full copy of data for respective component projects. Data shared with NERACOOS are backed up using the following protocols:

- **The data are protected using standard Linux security best practices including implementing firewalls, updating security patches, and restricting SSH users and hosts.**
- **The data are backed up on a regular basis (once per day) using AWS Cloud backup services.**
- **Copies of the data are stored on the cloud to mitigate against the event of a disaster.**

- 6.3. If there will be limitations to data access, how will these data be protected from unauthorized access? How will access permissions be managed? What process is to be followed in the event of unauthorized access?

N/A

7. Long-Term Archiving and Preservation

Note: NOAA's [Procedure for Scientific Records Appraisal and Archive Approval](#) describes how to identify, appraise and decide what scientific records are to be preserved in a NOAA archive.

- 7.1. In what NOAA Data Center (NODC, NCD, NGDC) will the data be archived and preserved? Have you begun discussions with that Data Center regarding your intended submission?

Observing data served through NERACOOS will be archived at NODC. The HF Radar DAC at NDBC will be responsible for archiving NERACOOS HF Radar data and the Glider DAC will be responsible for archiving glider data. Other NERACOOS datasets not mentioned above will not be archived. These other datasets are Federal datasets exposed by NERACOOS and are not at risk of being lost if NERACOOS does not archive them. Preliminary discussions with Steve Rutz at NODC took place in March 2014, and a Submission Information Form is currently being finalized with NODC. Archiving will likely begin by June 2014.

- 7.2. If you have not identified a NOAA Data Center, what is your long-term strategy for maintaining, curating, and archiving the data?

N/A

- 7.3. How will the costs of long-term data archiving be provided and maintained?

A portion of the funding provided through the cooperative agreement between the NOAA IOOS Program Office and NERACOOS will cover long-term data archiving.

- 7.4. What transformations or procedures will be necessary to prepare data for preservation or sharing? (e.g., quality control, format conversion, anonymization of personally-identifiable information, etc.). What related information will be submitted to the archive to enable future use and understanding of the data [e.g., metadata, references, reports, research papers, algorithms, audio or video codecs, special character sets or fonts, etc.] .

Because the data does not contain PII and all data is quality controlled, no additional requirements for archiving have been identified. Data for archiving will be submitted to NODC with all existing documentation, including metadata.

- 7.5. Identify the Record Schedule applicable to these data and provide the retention time for these data.

Observational data will tentatively be pulled by NODC from NERACOOS via THREDDS once per month. Archived data will be retained at NODC in perpetuity. These details are tentative and will be confirmed with the completion of the Submission Information Form. This data management plan will be updated to reflect the final archiving plan.