



CARICOOS

CARICOOS Data Archival Procedures

CARICOOS is working with the National Centers for Environmental Information (NCEI) in Silver Spring towards the long-term archival of applicable CARICOOS data holdings. The NCEI IOOS representative is advising us on the recently implemented, applicable, data submission procedures. Documentation of the archival process for our coastal data buoy network has been established in the Advanced Tracking and Resource tool for Archive Collections (ATRAC) system, project ID 8774. Attached is the Request to Archive documentation in its current state as of 2017-02-21 (Request2Archive_2017-02-17T18:29:04.pdf). CARICOOS has started the development archival procedures with NCEI for the WindNet data.

CARICOOS serves five data sets that already have an archive mechanism in place. These include: 1) wave data that are processed via CDIP, 2) glider data that are processed by the Glider DAC, 3) HFR data that are processed by the HFR DAC, 4) drifter data that are processed by AOML (the functional drifter DAC) and 5) meteo station data that are processed by WeatherFlow. Model output and data-derived products are not included in the archival process.

The Mesonet meteorological stations (currently numbering 13 as of September 2017) are owned and operated by WeatherFlow under contract to CARICOOS. Data from the station network are not submitted to NDBC and are not archived at NCEI under the current contractual agreement with CARICOOS. WeatherFlow data are collected at the station, stored in the station's data logger, and transmitted to the WeatherFlow central database (typically every five minutes). In the event the data cannot be transmitted (due to a loss of cell, Internet, or phone connectivity, etc.) the data are stored locally until transmitted, at which point all data collected since the last successful transmission is uploaded so no data is lost.

Observation data are retained permanently in the WeatherFlow database, and observation values are also pushed to Synoptic Data Corp., which also keeps a permanent archive.

Finally, all observations are pushed to the National Weather Service's (NWS) Meteorological Assimilation Data Ingest System (MADIS) by CARICOOS, where they are available for use within NOAA and NWS in near real time. These data are also stored and made freely available through two redundant CARICOOS THREDDs/OPeNDAP servers, one of which is located at an AWS instance and the other resides in a server

bunker at the University of Puerto Rico at Mayaguez.

Thus, the data that CARICOOS will be submitting for long-term archival to NCEI are therefore limited to: 1) oceanographic and atmospheric data from our coastal buoy network comprised of five UMaine buoys and 2) meteo station data that are processed by NDBC. The archive process is done via a Request to Archive procedure through NCEI's ATRAC system (<https://www.ncdc.noaa.gov/atrac/index.html>). CARICOOS has procured the fore mentioned ATRAC account for data archival purposes, and with Mr. Biddle's (IOOS Data Management Analyst) assistance this account has been configured accordingly.

UMaine in collaboration with NCEI is bringing the NetCDF buoy data files to full compliance with all applicable CF Conventions and Metadata standards (currently CF 1.6) and with NetCDF Attribute Conventions for Data Discovery (ACDD) through the use of IOOS and NCEI compliance checkers. UMaine's QC variables are also being updated through the addition of QARTOD flag conventions. In order to expedite the data archival process, it has been agreed upon that UMaine will directly submit data files from CARICOOS and NERACOOS to a secure data upload access point at UMaine for archival by NCEI. A secondary, or backup, secure data upload access point, enabled for cyclic redundancy checks of the files to be uploaded, has also been created in one of CARICOOS' data servers.

The site we have established (<http://dm2.caricoos.org/BROWSE/files/>) was tested by NCEI and declared as perfect for automation. NCEI was able to copy test files over and validate using the sha384 algorithm.

The Request to Archive documentation in CARICOOS's ATRAC account includes the following:

1. list of parameters/observations being collected.
2. processing steps/quality control including final format.
3. timing of data submissions and approximate sizes.
4. development of data documentation (metadata).
5. data disposition (path to archive center).
6. data affiliations, including both institutions and individual persons whose names will be associated with the data set in some way, e.g., where did it come from, where does it go, etc.

These items are outlined in the following sections for the two datasets we will be submitting for long-term archival to NCEI.

A. Coastal data buoy network

CARICOOS currently supports two different coastal data buoy types: five UMaine oceanographic data buoys. Dr. Neal Pettigrew of the Physical Oceanography Group at the University of Maine Ocean Observing System (UMOOS) is responsible for operating the UMaine oceanographic data buoys. This program is funded by IOOS through the CARICOOS. The data from the buoy system is managed by UMOOS and distributed to NDBC.

1. **List of parameters:** (see Section II.1 in the DMS Plan)
2. **Processing steps:** (see Section II.1 in the DMS Plan).
3. **Timing of submission:** UMaine generates near-real time NetCDF files for each buoy's current deployment (referred to as Realtime) and aggregated NetCDF files for past deployments (referred to as Historical). For archival purposes we propose submitting the updated Historical data files after the latest buoy maintenance cycle; these files will contain all buoy data up to the end of the last full deployment. The Historical data files would be updated annually upon completion of the buoy maintenance cycle, usually in June-July but may vary.
4. **Data documentation:** Submitted NetCDF files contain metadata records of buoy location, instrumentation, points of contact, QC flags, etc.. Other documentation is provided on the CARICOOS web site (<http://caricoos.org>).
5. **Data disposition:** The NCEI approved secure data upload access points at UMaine (URL to be provided by UMaine) and at CARICOOS <http://dm2.caricoos.org/BROWSE/files/>, which are enabled for cyclic redundancy checks of the files to be uploaded by NCEI.
6. **Data affiliations:**
 - *Submitting institution:* CARICOOS
 - *Data point of contact:* Dr. Patricia Chardón-Maldonado (patricia.chardon@upr.edu)
 - *Instrument technical point of contact:* Dr. Neal Pettigrew (nealp@maine.edu)
 - *Data technical points of contact:* Mr. Bob Fleming (bffleming@umeoce.maine.edu)

and Dr. Patricia Chardón-Maldonado (patricia.chardon@upr.edu)

7. **Schedule:** A conversation has been initiated with NCEI and UMaine and initial transfer of the data collected can happen as early as the 3rd calendar quarter of 2017

B. WindNet network of meteorological observations

In addition to the WeatherFlow--operated Mesonet CARICOOS currently operates and maintains two land--based meteorological stations in its WindNet.

Dr. Patricia Chardón-Maldonado of CARICOOS is responsible for operating the meteorological stations. The data from both stations are acquired by CARICOOS and distributed to NDBC for near--realtime processing.

1. **List of parameters:** (see Section II.4 in the DMS Plan)
2. **Processing steps:** (see Section II.4 in the DMS Plan).
3. **Timing of submission:** Near--realtime NetCDF files for each station are generated at 15-minute intervals and are updated at the **Data disposition** URL accordingly. A single NetCDF containing all aggregated data will be submitted for each station; current file sizes are smaller than 3 Mb.
4. **Data documentation:** Submitted NetCDF files contain metadata records of station/sensor location, instrumentation, points of contact, QC flags, etc.. Other documentation is provided on the CARICOOS web site (<http://CARICOOS.org>).
5. **Data disposition:** The NCEI approved a secure data upload access point at CARICOOS <http://dm2.caricoos.org/BROWSE/files/>, which is enabled for cyclic redundancy checks of the files to be uploaded by NCEI; i.e., a CARICOOS push / NCEI pull configuration.
6. **Data affiliations:**
 - *Submitting institution:* CARICOOS
 - *Data point of contact:* Mr. Jose Torres (jose.torres111@upr.edu)
 - *Instrument technical point of contact:* Dr. Patricia Chardon-Maldonado (patricia.chardon@upr.edu)
 - *Data technical points of contact:* Mr. Jose Torres(jose.torres111@upr.edu)

7. **Schedule:** A conversation has been initiated with NCEI and initial transfer of the data collected can happen as early as the 3rd calendar quarter of 2017.