CARICOOS
Data Management System (DMS) Plan

I. Executive Summary

This document provides an overview of one of the major components of CARICOOS, as required by the NOAA IOOS – the Data Management System (DMS). The CARICOOS DMS currently operates as the Caribbean Regional Data Assembly Center (DAC). The fundamental function of the DMS is to aggregate multiple data streams from the sensors and models that comprise CARICOOS as well as from independent data providers into a central archive and provide these data to users via standard services. The CARICOOS DMS is involved with all aspects of data flow including archive, discovery, and transport with efforts primarily geared toward: 1) obtaining and distributing a variety of quality data from external partners; 2) managing model data output (ROMS, WRF, SWAN, AMSEAS); 3) maintaining the flow of data into the Data Access Services and web products; and 4) enhancing product development. It archives and serves data and model output for the entire U.S. Caribbean geographical region and facilitates discovery, access, and understanding of regional, relevant in-situ and model data. CARICOOS data servers are registered and listed in the IOOS catalog, allowing them to be discovered. Data can be accessed via WMS, THREDDS/OpenDAP, ncSOS, ERDDAP, and visualized in the CARICOOS portal (http://www.caricoos.org/). Data may also be provided to stakeholders in several standard formats (such as csv, xls, txt, mat or NetCDF) upon request.

CARICOOS has been developing the DMS for archiving and serving data and model output over the past several years with grants from NOAA IOOS. These efforts require substantial information technology support in the form of system administration of the various computers and networking between instruments, models, and the data server systems.

II. System Components / Data Streams
1. UMaine Coastal Buoys / Meteorological and Oceanographic
2. UVI A Coastal Buoy (VIA) / Meteorological and Oceanographic
3. Mesonet Weather Stations / Meteorological
4. WindNet / Meteorological
5. Rincon Wave Buoy / Surface Wave
6. High Frequency Radar (HFR) / Surface Water Velocity
7. MAPCO2 Buoy / Atmospheric and Water Chemistry
8. Gliders / Conductivity-Temperature-Depth Profiles
9. Drifters

III. Data Servers, Services and Flow / Common to Many Data Streams

IV. General Comments on Data Quality Assurance and Control, and Data Archiving
1. Quality Assurance
2. Quality Control
3. Data Archiving
The DMS Plan focuses on the management and delivery of CARICOOS-related data. CARICOOS will implement recommended and standard practices as defined by the IOOS Data Management and Communications (DMAC) committee and more specifically those in the Guide for IOOS Data Providers, version 1.0 (2006). These practices apply to data archive, data discovery, data service (web-based browsing), data transport (access to data), metadata, IT security, and data QA/QC.

CARICOOS adheres to the NOAA Data Sharing Procedural Directive. All real-time data collected by CARICOOS are freely available through open services, without delay or restriction. Avenues for accessing the data are available through the CARICOOS website: http://www.caricoos.org/data-download. At present CARICOOS does not maintain any data streams that are restricted, either to specific users or after delays.

Whenever guidance is provided by the U.S. IOOS Program Office on data management protocols, CARICOOS will respond within 1 week 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, he takes the necessary steps towards its implementation in a reasonable and timely manner. Implementation of new services is only limited by personnel time and expertise. CARICOOS often consults outside DMAC expertise from our sister RAs (or RICEs). Additionally, CARICOOS DMS personnel maintain regular communication with the U.S. IOOS Program Office through in-person meetings, phone calls, webinars, and emails. This continuous communication ensures that CARICOOS is aware of all new practices and protocols, as promulgated by the IOOC and the IOOS Program Office, and understands how to implement them.

CARICOOS seeks to comply with IOOS-DMAC data server and services specifications provided in

DMAC Implementation Plan

The latest version of the DMAC Plan (2011), describing DMAC requirements, architecture, planned implementation, etc can be found here: DMAC Plan 1.0 (.pdf)

IOOS DMAC Subsystem Implementation Guidance

NOAA IOOS® Program Office White Paper designed to advance community discussion about the national/regional IOOS® enterprise, specifically with a focus on implementation of DMAC subsystem elements. IOOS DMAC Subsystem Implementation Guidance

Concept of Operations for the DMAC subsystem
The initial high-level concept of operations (ConOps) for the DMAC subsystem: IOOS DMAC Concept of Operation (.pdf)

IOOS DMAC FAQ

The live DMAC question and answer web page can be found here: IOOS DMAC FAQ (.pdf)

II. System Components / Data Streams

CARICOOS is a unique Regional Association in that much of its data collection, operations, and data management are conducted through contracting arrangements. Individuals and/or organizations are sub-contracted by CARICOOS to provide operations and maintenance for the various system components of CARICOOS. There is a system in place to check and evaluate the contracting agreements. Contractors provide semi-annual reports to the University of Puerto Rico – Mayaguez (UPRM) to demonstrate that they are meeting the agreed-upon scope of work. Both the Executive and Technical Director review the reports and approve disbursements to the contractors based on the services rendered. Additionally, CARICOOS staff perform daily checks to assure all data streams are active. Uptime and response time in case of interrupted data streams are criteria weighted in the performance assessment of the data-serving contractors.

This section describes each CARICOOS data source individually. The descriptions include the data flow and quality control procedures. Existing or planned archiving procedures for each data stream are described in a separate document.

1. UMaine Coastal Buoys / Meteorological and Oceanographic

Dr. Neal Pettigrew’s lab at the University of Maine operates and maintains 4 buoys for CARICOOS (PR1=Ponce; PR2=San Juan; PR3=Vieques; VI1=St. John). The standard measurements on these buoys include near-surface winds (Gill Windsonic anemometer and RMYoung propeller-vane wind sensor), air temperature (Campbell Scientific air temperature), barometric pressure (Setra pressure sensor), directional waves (Axys Triaxys wave sensor), ocean surface temperature and salinity (Seabird 37SM), and ocean currents throughout the water column (Nortek Aquadopp current profiler). An in-house (UMaine-built) directional wave sensor is installed on PR3 (no Axys Technologies sensor); the same sensor is installed on the other buoys as a redundant directional wave sensor. A third non-directional wave sensor (based on single axis accelerometer data) is installed on all buoys as a redundant data source and comparison check. Wave data reported to NDBC are from the Ayxs sensor if available, and from the UMaine-built directional wave sensor if not.

Data Flow
Under contract to CARICOOS, UMaine acquires the buoy data through two redundant communication systems: 1) mobile cellular service (primary) and 2) satellite transmission (backup). The data from all instruments are connected to a Campbell Scientific CR1000 data logger in each buoy for collection and transmission via cellphone IP data transfers. Backup transmissions are also received via NOAA’s GOES satellite systems. The data streams are ingested into their system, where all real-time data are processed and quality controlled by UMaine, and distributed to CARICOOS and to NDBC. Note that the data loggers in the buoys transmit at hourly intervals so buoy data are updated hourly, despite the shorter measurement time step for some parameters.

CARICOOS accesses UMaine buoy data in 3 ways:

a. Through the CARICOOS Buoy Monitoring web page that was set for us by UMaine (http://gyre.umeoce.maine.edu/caricoos/). The last 10 days of data from each buoy are provided in text format. Meteorological parameters (labeled “Surface”) at the height indicated for the corresponding station sensor are reported at 10 minute intervals (wind direction, wind speed, wind gust, atmospheric pressure, atmospheric pressure tendency and air temperature) whereas sea surface parameters (labeled “Ocean”) at the depth indicated for the corresponding sensor are reported at 1 hour intervals (water temperature, salinity, wave height, wave direction, wave period, surface current speed and surface current direction). Water velocity profiles (labeled “Currents”) are reported at 1 hour intervals. These are data listings with no imbedded metadata records. These data have been submitted to UMaine’s standard QARTOD procedures. Note data logger transmission caveat above.

b. UMaine provides historical (up to the previous buoy deployment) and near realtime (current buoy deployment) data in NetCDF format. These files are downloaded from http://gyre.umeoce.maine.edu/GoMoos/php/mooring_file_info.php?report=historical_files and http://gyre.umeoce.maine.edu/GoMoos/php/mooring_file_info.php?report=active_files, run through Python routines by CARICOOS to ensure NCSOS compatibility and are then uploaded to the THREDDS server.; these files contain extensive metadata records and QA/QC flags.

c. From NDBC in csv and NetCDF formats.

The data processing system for data generated by the UMaine buoys consists of both field and shoreside components. A programmable datalogger inside the buoy well timestamps data received from sensors, aggregates it, and transmits the data to UMaine via two methods, cellphone modem and NOAA GOES satellite. This results in two streams of largely redundant data.
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 are stored 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.

In the text-based data feed to CARICOOS, data from the two redundant data streams (GOES and cellphone) as well as data from redundant sensors (wind and wave data) are combined and reconciled. QC is carried out on both data streams independently as the data arrive. The text (ASCII) reports are updated as incoming data are processed, and may be “backfilled” as missing data or data sampled at a higher temporal resolution become available later in time. For example, in the event of a temporary cellphone service interruption, wind speed data sampled at hourly intervals may arrive via the GOES data stream, for a period of time, before wind data sampled at a 10-minute interval can arrive via the cellphone data stream to “fill in the gaps”.

As buoy data are processed and updated, associated NetCDF files are immediately replicated from the application server to the web server. Both servers are backed up nightly, and full backups of the NetCDF archive are retained off-premises periodically.


Quality Control

UMaine performs data stream QA/QC according to best practices and standards. All data variables reported to CARICOOS in realtime NetCDF files have an ancillary QC flag. A non-zero flag represents an invalid data value for that record. 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 for this purpose at an early stage in processing, and QC flags are set to a non-zero value. The ranges employed for testing are selected by buoy personnel and 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 realtime automated QC checks, data are reviewed by data management personnel on a daily basis using a variety of diagnostic tools. This operator-supervised data review includes the comparison of buoy data to data from neighboring platforms and to climatology data, serving a similar function to the QARTOD “Neighbor Test” and “Climatology Test”. A number of email alarms are configured to indicate gaps in data, unusual or anomalous sensor data, or system processing issues and notify data management personnel within a timeframe that minimizes downtime or gaps in data. Additional automated and operator-supervised QC is carried out on when sensors with onboard storage are recovered post deployment.

Automated QC checks for the in-situ current measurements (Nortek ADCPs) are limited and are targeted for improvement. UMaine is working to implement In-Situ Current QARTOD guidelines (released in October 2015) to the extent possible using current telemetry and instrument configuration by the late 3\textsuperscript{rd} quarter or 4\textsuperscript{th} quarter calendar 2017.

Status of required QC tests for the data variables listed in the QARTOD manuals are listed in Table 1. Data that fail a test are flagged with a non-zero QC flag.

UMaine data management personnel certify that as of October 2016, QC is being applied at a level equal to or greater than QARTOD standards, given that:

a) Most of the required QARTOD checks are already in place. (Please see Table 1)

b) Daily operator-supervised manual review of data supplements the QARTOD checks in place as described above.

c) A number of in-house automated QC checks for data other than those described in QARTOD standards have been developed over time and are currently in use. Examples include comparison of barometric pressure standard deviation with wind speed, comparison of data from redundant sensors, examination of minimum wind speeds and wind sample counts, and gradient checks for air temperature and barometric pressure.

We have the following plan for more complete implementation of QARTOD standards by the late 3\textsuperscript{rd} quarter or 4\textsuperscript{th} quarter calendar 2017.

a) Enhance realtime QC of Nortek ADCP data by implementing Sensor Tilt QC test, and other tests based on orientation data, Error byte, and Status byte;

b) Implement a gridded model-based comparison check for Wind, Air Temperature, Barometric Pressure, based on NOAA RTMA or Rapid Refresh short term forecasts for Puerto Rico;

c) Implement climatology tests for QARTOD (and other) variables based historical observations we have on hand and a “storm recognition” algorithm to recognize extreme events;
d) Implementation of a “1-based” set of QC flags for buoy data. This will be implemented in parallel with and in addition to the existing zero-based QC flagging field in the existing NetCDF files;

e) Update QA plan and QC process documentation.

Data Sharing

CARICOOS provides buoy data to clients:

a. CARICOOS THREDDS/OPeNDAP servers
(http://dm2.CARICOOS.org/thredds/catalog.html, and http://dm1.CARICOOS.org/thredds/catalog.html) provide whole NetCDF files and subset queries. Cron jobs in the THREDDS server are scheduled to download the Historical and Realtime NetCDF data files from UMaine, run ncSOS compliance scripts and aggregate the buoy data.

b. CARICOOS Data Download web page http://www.CARICOOS.org/data-download feeds from NDBC and UMaine csv data files

c. A dedicated CARICOOS web page for each buoy may be accessed through
http://www.CARICOOS.org/#!detail=SelectBuoy

CARICOOS provides buoy products to clients:

a. The map at http://www.CARICOOS.org the main entryway to the CARICOOS web portal, feeds from the 10-day data through an intermediate MySQL database

b. A dedicated CARICOOS web page for each buoy may be accessed
http://www.CARICOOS.org/#!detail=SelectBuoy station metadata, most recent data, time series plots and data download links are available from these pages
<table>
<thead>
<tr>
<th>In-situ temperature and Salinity</th>
<th>Temperature</th>
<th>Salinity</th>
<th>Conductivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1 Timing/Gap test</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Test 2 Syntax Test</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Test 3 Location Test</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Test 4 Gross Range</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Test 5 Climatology Test</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Test 7 Rate of change test</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wind Data</th>
<th>Wind Speed</th>
<th>Wind Direction</th>
<th>Wind Gust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1 Timing/Gap test</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Test 2 Syntax Test</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Test 3 Location Test</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Test 4 Gross Range</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Test 5 Climatology Test</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>In-situ Surface Wave Data</th>
<th>Significant Wave Height</th>
<th>Dominant Wave Period</th>
<th>Avg Wave Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 16 Stuck Sensor</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Test 17 Operational Frequency</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Test 18 LF Energy</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Test 19 Bulk Wave Params</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Test 20 Rate of Change</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>In-situ Current Observations</th>
<th>Current Speed</th>
<th>Current Direction</th>
<th>Current u</th>
<th>Current v</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 2 Check Sum</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Test 3 Sensor Tilt</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Test 4 Speed of Sound</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Test 6 Signal Strength</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Test 10 Current Speed</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Test 11 Current Direction</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Test 12 Horizontal Velocity</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Test 15 Stuck Sensor</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Test 16 Echo Intensity</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Table 1. UMaine QC tests for buoy data as of November 2016. Applicable QARTOD tests and NetCDF flags will be added in parallel starting by the late 3rd quarter or 4th quarter calendar 2017.
2. UVI A Ocean Buoy (VIA) / Meteorological and Oceanographic

UVI subcontracts with Caribbean Wind LLC (CWLLC) to provide regular maintenance and data management. The UVI A Ocean Buoy (currently at Brewer’s Bay, St. Thomas but may be relocated by UVI) measures the following parameters: significant wave height, maximum wave height, dominant wave period, mean wave direction, mean wave direction spread, air temperature, relative humidity, atmospheric pressure, wind direction, wind speed, wind gust, water temperature, salinity, dissolved oxygen, sigma_t, average chlorophyll concentration, sea water turbidity and water velocity profiles starting at a depth of 2 meters.

Presently data are provided directly to the University of Maine (UMaine) for processing and distribution. Note that once UMaine ingests the data from the CARICOOS UVI A Buoy, the data are treated identically to the data from the other CARICOOS/UMaine buoys, as described above.

The data flow, quality control procedures, data sharing and archival for the UVIA buoy mirror that of the other CARICOOS/UMaine buoys, as described above.

http://gyre.umeoce.maine.edu/caricoos/

3. Mesonet Weather Stations / Meteorological

WeatherFlow owns and operates 14 Coastal Mesonet weather stations for CARICOOS. The standard measurements for these stations include surface winds (R.M. Young wind sensor), air temperature, and barometric pressure. Under contract to CARICOOS, Weatherflow installs, maintains and collects the data from the Mesonet network of meteorological stations.

Data Flow

The data processing system for data generated by the WeatherFlow weather stations for CARICOOS consists of both field components and shoreside components. Programmable dataloggers, built by WeatherFlow, are mounted on various locations. The dataloggers timestamp data received from sensors, aggregate it, and transmit it to WeatherFlow via GPRS radio signal. A brief description of the data flow can be found here: WF Data and Coms

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 are stored and archived in NetCDF format, following CF metadata and COARDS NetCDF conventions in effect when the system was built. Additional metadata for NetCDF files and directory structure are accessed via MySQL databases.

Data are transferred to CARICOOS after initial QA/QC; Weatherflow retains ownership of the hardware.
a. The last 24 hours of data from each station are provided in text format at http://datascope.weatherflow.com/externalDataView.cfm?special=UPR2. Wind direction, atmospheric pressure, air temperature and mean, lull and gust speeds are included.

b. These 24-hour data are ingested into the MySQL database and from here to the CARICOOS.org map (http://www.caricoos.org) and products. They are also used to create the NetCDF realtime files for the THREDDS/OPeNDAP servers. CARICOOS operates duplicate THREDDS servers for Mesonet data at Amazon Web Services (http://dm1.caricoos.org/thredds/catalog/content/Mesonet/catalog.html) and at UPRM-Mayaguez facilities (http://dm2.caricoos.org/thredds/catalog/content/Mesonet/catalog.html); note these servers are located at different geographical locations.

c. Once a month data from the previous month are manually downloaded from Weatherflow’s DataScope server, plotted and visually checked. Monthly NetCDF are then created and added to the THREDDS/OPeNDAP servers, and the realtime files are reset.

d. Data are submitted to the GTS system by CARICOOS.

e. Some old historical csv data files are available through Data Download but the main distribution source is our THREDDS/OPeNDAP server.

f. Mesonet data are not ingested by NDBC and are not archived at NCEI due to restrictions in the contractual agreement between WeatherFlow and CarICOOS.

Quality Control

WeatherFlow currently conducts a set of QC checks on all of its operational data (wind speed and direction, air temperature, and air pressure). The full scope of the details of this collection of tests is proprietary, but all include a Timing/Gap check, a Syntax check, a Location test, and a Gross Range test. WeatherFlow QC procedures are equivalent to or greater than the minimum standards identified by QARTOD.

Data Sharing:

CARICOOS provides Mesonet data and products to its clients through its publicly accessible OPeNDAP/THREDDS data servers.

4. WindNet / Meteorological

Dr. Patricia Chardon of CARICOOS operates and maintains two WindNet land-based coastal weather stations for CARICOOS. The standard measurements for these stations include wind speed, barometric pressure, and air temperature. The NDBC station codes, PTRP4 for Rincon and IMG4 for Isla Magueyes, are used through the data flow / data processing pipeline.

Data Flow

WindNet data flow proceeds as follows:
a. CARICOOS acquires the station data via internet through mobile cellular service or through direct internet connection depending on a station’s location. Both WindNet stations are equipped with Sutron Xlite 9210B data loggers.

b. Near-realtime data are disseminated from a Caricoos server to NDBC via ftp following NDBC’s data messaging protocols.

c. NDBC sends the data to GTS, performs QA/QC and makes these data available through near-realtime and historical csv data files, and also in NetCDF format through its THREDDS/OPeNDAP server.

d. CARICOOS pulls the NDBC near-realtime csv data and generates NetCDF files that are distributed through our THREDDS servers. Cron jobs are scheduled in our DMAC servers to download the NDBC csv files, generate aggregated NetCDF files with augmented metadata and upload these aggregated data files to the CARICOOS THREDDS/OPeNDAP servers. The data are made available through our duplicate servers.

e. Our web products use the NDBC csv data.

Quality Control

CARICOOS staff procedures and practices comply with Appendix A of the QARTOD manual in the commission of the current stations (PTRP4 & IMGP4). Both WindNet stations are equipped with Sutron Xlite 9210B data logger. The data logger performs an internal QC check to flag the average data. The quality can be G=GOOD, B=BAD or U=UNDEFINED. A quality status of UNDEFINED means that the system has not yet tried to measure the sensor. Data with B or U are removed; the other data are disseminated to NDBC.

NDBC performs full QA/QC on the data provided by WindNet.

Data Sharing

NDBC web links:


CARICOOS THREDDS server links:

- [http://dm1.caricoos.org/thredds/catalog/content/WindNet/catalog.html?dataset=windnet/ptrp4_realtime_ndbc_qc.nc](http://dm1.caricoos.org/thredds/catalog/content/WindNet/catalog.html?dataset=windnet/ptrp4_realtime_ndbc_qc.nc)
- [http://dm2.caricoos.org/thredds/catalog/content/WindNet/catalog.html?dataset=windnet/imgp4_realtime_ndbc_qc.nc](http://dm2.caricoos.org/thredds/catalog/content/WindNet/catalog.html?dataset=windnet/imgp4_realtime_ndbc_qc.nc)
5. Rincon Datawell WaveRider Buoy / Surface Wave

The Datawell WaveRider buoys (1 operational, 1 spare) are owned and maintained by CARICOOS. The Coastal Data Information Program (CDIP) provides data management services. The parameters measured include wave height, wave direction, wave period, and water temperature.

Data Flow

a. The Waverider buoy transmits directly to CDIP via the Iridium satellite constellation
b. CDIP sends the data to NDBC, both CDIP and NDBC QC the data stream
c. CARICOOS web products and data download options use the 45-day csv data from NDBC
d. The CARICOOS THREDDS carries copies of the CDIP historical and real-time NetCDF data files.

Quality Control

CDIP handles the acquisition and dissemination of the WaveRider buoy data from UC San Diego-SCRIPPS. CDIP performs QA/QC and archival procedures on all data collected by the Rincon Waverider Buoy.

Data Sharing

The CARICOOS THREDDS servers provide copies of the CDIP NetCDF data files to all stakeholders at:

http://dm1.caricoos.org/thredds/catalog/content/Rincon_Waverider/catalog.html

http://dm2.caricoos.org/thredds/catalog/content/Rincon_Waverider/catalog.html

6. High Frequency Radar (HFR) / Surface Water Velocity

HFR technology provides real-time surface current velocities. There are 5 HFR antennas in the CARICOOS region. CARICOOS owns and operates two (South Cabo Rojo and Rincon) HFR. CARICOOS also operates the North Cabo Rojo HFR (owned by Texas A&M) and the Ponce and Maunabo HFRs (leased from Rutgers). HFR current maps for the Mona Passage are made possible through support from the Department of Homeland Security to the Center for Secure and Resilient Maritime Commerce and Coastal Environments (CSR). Technical and administrative support to CSR is provided by CARICOOS in the operation of 2 transmit/receive antenna pairs located at Club Deportivo del Oeste in Cabo Rojo and the Puerto Rico Police FURA station on
Añasco beach. CARICOOS currently provides coverage for the west coast, southwestern coast, and much of the southern coast (including the ports of Ponce and Guayanilla) of Puerto Rico.

**Data Flow, Quality Control**

a. Data from the 5 HFR antennae are treated the same. The antenna raw data are sent via wireless internet to the CARICOOS manoa server at Isla Magueyes. Rutgers and the IOOS HF Radar DAC pull the raw data from manoa and process, QA/QC, archive and distribute them.

b. The HFR DAC consists of developmental servers at Scripps which also serve as the primary operational server. Backup operational servers are located at NDBC, and at Rutgers for fail-over security.

c. CARICOOS has full data archives and processes new data at hourly intervals (for internal use only at the moment).

**Data Sharing**

a. Once the HFR radial data are quality controlled, total vectors are calculated and displayed on the National HFR Network, hosted by Scripps, CORDC: ([http://cordc.ucsd.edu/projects/mapping/maps/](http://cordc.ucsd.edu/projects/mapping/maps/))

b. Data are made available via the CORDC THREDDS Server ([http://hfrnet.ucsd.edu/thredds/catalog.html](http://hfrnet.ucsd.edu/thredds/catalog.html)) and the NDBC OpeNDAP/THREDDS Server ([http://sdf.ndbc.noaa.gov/thredds/catalog.html](http://sdf.ndbc.noaa.gov/thredds/catalog.html))

c. Links to Scripps (CORDC) graphical web products for the Caribbean are provided in the CARICOOS web page (see [http://www.caricoos.org/hfradar](http://www.caricoos.org/hfradar)).

d. Scripps (CORDC) total vectors are also re-plotted and animated locally. The vector movies for our region are found at [http://www.caricoos.org/currents/observation/hfradar/regional/6km](http://www.caricoos.org/currents/observation/hfradar/regional/6km)

7. **MAPCO2 Buoy / Atmospheric and Water Chemistry**

NOAA, Pacific Marine Environmental Laboratory (PMEL) provides operations and data management for the MAPCO2 buoys in the CARICOOS region. The MAPCO2 buoys measures atmospheric CO2 concentration, water CO2 concentration, water pH, salinity, and sea surface temperature.

a. Collaboration with NOAA, Pacific Marine Environmental Labs

b. Data available from CDIAC ([http://cdiac.ornl.gov/oceans/Moorings/La_Parguera.html](http://cdiac.ornl.gov/oceans/Moorings/La_Parguera.html)) and from PMEL ([http://www.pmel.noaa.gov/co2/](http://www.pmel.noaa.gov/co2/))

c. CARICOOS downloads PMEL data and constructs trend plots that are displayed in [http://www.caricoos.org/oceans/acidification/seawater](http://www.caricoos.org/oceans/acidification/seawater)
d. PMEL performs all QA/QC on data from the MAPCO2 Buoy. Details of general instrument QA/QC process for each deployment are found at: http://cdiac.ornl.gov/oceans/Moorings/La_Parguera.html.

e. Preliminary data are transmitted to PMEL daily by the iridium satellite data transmission system. Junk data are flagged and removed by PMEL. An internal QC process for atmospheric CO2 concentration, water CO2 concentration, and water pH is conducted by CARICOOS. This process consists in a matlab subroutine that updates daily and identifies values outside of ±3 standard deviations of the mean to subsequently post it at CARICOOS website at http://www.caricoos.org/oceans/acidification/seawater.

f. PMEL preliminary data are submitted to the Carbon Dioxide Information Analysis Center (CDIAC) for final QA/QC process and archival. This process follows the CDIAC standard QA/QC process (more details at http://cdiac.ornl.gov/ftp/cdiac74/chapter3.pdf).

Metadata are provided by:

- Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee. doi: 10.3334/CDIAC/OTG.TSM_La_Parguera_67W_18N.

8. Gliders / Conductivity-Temperature-Depth Profiles

The NOAA, Atlantic Oceanographic & Meteorological Laboratory (AOML) owns and provides operations and data management for several gliders in the CARICOOS region (four as of November 2016). CARICOOS provides field support to the gliders in the form of ship time for deployments and retrievals plus local students and crews in collaboration with AOML. The gliders measure vertical profiles of temperature and salinity.

a. Glider data are received by AOML and forwarded to the Glider DAC who performs the appropriate data management and QC procedures.

b. “Glider DAC: In 2013 the U.S. IOOS Program Office established a prototype national glider DAC, developed by MARACOOS, as a central access point for this rapidly emerging technology. The glider DAC’s initial rapid deployment was possible because IOOS has focused on data standards, the large number of glider missions flown by IOOS and academic partners, and leveraging of other platforms data standards, e.g. ARGO. (http://gliders.ioos.us)”. From http://www.iooc.us/ioos-dmac-frequently-asked-questions.
c. Data and products are available through AOML. CARICOOS serves glider data via the AOML link found in our glider web pages (see below):
   http://www.aoml.noaa.gov/phod/goos/gliders/observations.php.

   d. CARICOOS serves value-added glider products generated in-house in its webpage http://www.CARICOOS.org/gliders.

9. Drifters
The NOAA, Atlantic Oceanographic & Meteorological Laboratory (AOML) owns and provides operations and data management for the Global Drifter Program surface drifters in the CARICOOS region. The drifters measure horizontal trajectories.

   a. CARICOOS provides field support to our disposable drifters in the form of ship time for deployments plus local students and crews in collaboration with AOML.
   
   b. Data and products are available through AOML. CARICOOS serves GDP drifter data through its Data Download webpage via the AOML link: http://www.aoml.noaa.gov/phod/dac/gdp_doc.php.
   
   c. AOML provides all QA/QC for CARICOOS drifters: www.aoml.noaa.gov/phod/dac/Training_CD.pdf.

10. Citizen Science Data / Beach Water Quality

Surfrider Water Quality Data.

Water quality tests for beach water Enterococcus and bacteria concentrations in Puerto Rico and the US Virgin Islands are performed by the Surfrider Foundation-Rincon, by the Puerto Rico Environmental Quality Board, and by the CARICOOS team following EPA-approved Enterolert® QA protocols. The results show the Most Probable Number (MPN) of colony-forming units (CFU) of Enterococcus in 100ml of seawater. These data are at least two days old by the time they reach CARICOOS. The equipment and protocols used by Surfrider are included in the Standard Operating Procedures document.

Data Flow
CARICOOS grabs the data directly from the Surfrider Blue Water Task Force site and from the Environmental Quality Board via api. Enterococcus data and time series plots are made available through CARICOOS.

Quality Control

No quality control procedures are performed on these data so user discretion is advised. CARICOOS provides a disclaimer on these data.

Data Sharing
The geographical distribution of sampled beaches is displayed in the CARICOOS webpage http://www.CARICOOS.org/map/beach-water-quality. Through the interactive map a user may quickly scan the levels of enterococcus contamination in the region and also access the most recent concentration values as well as past time series at each sampled location.

III. Data Servers, Services and Flow / Common to Many Data Streams

The backbone of the CARICOOS data system is the data service. On the back-end of this system is an OPeNDAP-based architecture. In this section, the data transport and discovery tools are described.

The data system can be envisioned as a tiered system, with data at one end and client tools, including web page displays, at the other. Connecting these two are databases and file systems, data servers, and data services. Figure 1 shows this schematically.

CARICOOS maintains duplicate/redundant THREDDS/OPeNDAP servers: 1) dm1.caricoos.org is housed in an instance server at Amazon Web Services (AWS, in the Cloud) and 2) dm2.caricoos.org is housed at the UPR–Mayaguez campus server bunker. These two server facilities are geographically separated by over a thousand miles and constitute a duplicate/redundant pair in terms of their data holdings (but may differ in terms of model output). Users may access either server for their data needs.

The THREDDS server hardware at UPR consists of generic rack-mounted, multi-processor, multi-core linux systems with ample storage space on which the appropriate Java-Apache-THREDDS-OPenDAP software scaffold is installed. Software versions and upgrades are recommended by IOOS and Unidata while several github locales and the ioos_tech usergroup provide a certain degree of technical support. AWS provides virtual server instances tailored to our hardware and operating system specifications; server OS configuration and software installation are performed by CARICOOS.

All data accessible via the THREDDS servers resides within local storage in each server. Scripts running at periodic intervals in the servers (aka cron jobs) manage the acquisition of observational data from the various server providers and modeling output from either local high performance computing servers or from community-model output repositories. The data fetching process has been tuned to balance optimal data latency versus network access limitations.

Our ArcGIS Server resides at Isla Magueys and feeds from several ArcGIS front-end production systems. Initially configured to serve geospatially referenced storm surge inundation information it has also been used for satellite imagery products.
On-site storage is distributed among our various servers and dedicated NAS. Most of the storage space is dedicated to model output. AWS S3 storage buckets provide additional storage space in the Cloud and are used for a variety of purposes.

All data served by CARICOOS moves through to the CARICOOS THREDDS server and may or may not be available through all the data services (WMS/WCS, ncSOS, OPeNDAP, HTTP). Not all data are suited for each of the four data services CARICOOS provides. From there, the data are made available to viewers and clients through the following platforms:

- **Web Portal**
  - The CARICOOS web portal, [http://www.caricoos.org](http://www.caricoos.org), serves as the main window for viewing and accessing all data served.

- **CARICOOS Explorer**
  - The CARICOOS Explorer is a user-friendly graphical interface that combines model output with observations and forecasts. It allows the user to explore various data levels at a time. [http://assets.maracoos.org/?config=cari](http://assets.maracoos.org/?config=cari)

- **ERDDAP**
  - ERDDAP is an interface for the graphical output of model data in 2D and 3D. ERDDAP runs on the OPeNDAP servers.

- **Data Download**
  - User-friendly method for users to download entire datasets. Datasets are provided in ASCII format for ease of use. [http://www.caricoos.org/data-download](http://www.caricoos.org/data-download)

CARICOOS migrated its database, one OPeNDAP/THREDDS, the scripts server and several computational servers to the Cloud in the third quarter of FY 2016-2017. Our web page database, the scripts server and the OPeNDAP/THREDDS server are operational as Amazon Web Services (AWS) EC2 containers while the ftp server was reconfigured as AWS S3 buckets.
IV. **General Comments on Data Quality Assurance and Control, and Data Archiving**

With the exception of a couple assets, all of the data served through CARICOOS are provided by external sources, including subcontracted organizations, partners, and Federal data servers.

CARICOOS serves observational Near Real-Time Data which we define as being transferred from a sensor package to the respective asset providers at the same frequency the data are collected, with minimal latency for any given observation, where latency is being defined as the time interval between the moment an observation is taken and the moment when it becomes available to our stakeholders. The asset provider performs QA/QC on these data and they become available for download into the CARICOOS data servers. Our goal is to minimize latency times to the full extent as possible.

CARICOOS manages Citizen Science data from Surfrider. These water quality data are collected by members of the general public who are not trained scientists. These data are collected by volunteers and have limited quality control.

1. **Quality Assurance**

CARICOOS relies on subcontracted partners and established partner programs (i.e. CDIP) to provide the best possible QA/QC on instruments. Each asset provider operates and maintains their equipment in compliance with manufacturer guidelines (see CARICOOS Equipment Standard Operating Procedures and Inventory document).
CARICOOS serves near real-time data provided by federal agencies, IOOS Data Assembly Centers (e.g. CDIP, HFR, gliders), and external groups (WeatherFlow, UMaine). These entities perform their own QA/QC according to best practices and standards. CARICOOS does not serve near real-time data from external partners who do not perform QA/QC.

2. Quality Control

CARICOOS does not apply near real-time quality control procedures to the data it distributes, but all data made available by CarICOOS are quality controlled by the data originators.

All of the near real-time data received and served by CARICOOS are quality controlled. The quality control methods vary with each provider. However, each provider is contractually required to implement QARTOD procedures (if applicable) or provide QA/QC at a level equal to or greater than QARTOD standards.

All data available through CARICOOS are either provided by:

Federal Agencies:

- MAPCO2 buoys
- Drifters
- Data from the WindNet stations, operated by CARICOOS, are ingested and served through NDBC, and run through the NDBC QC procedures.
- NWS Doppler Radar
- NASA Ocean Color
- NOAA-NOS tide gauge

CARICOOS, via a functioning National Data Assembly Center (DAC) who performs QA/QC:

- HFR
- Wave Buoy
- Gliders

By a partner with documented QA/QC procedures:

- UMaine buoys
- UVI-A ocean buoy
- Weatherflow meteo stations

3. Data Archiving

The CARICOOS Data Archiving Plan is included in a separate document.

CARICOOS has initiated the data archival process with NCEI through Mathew Biddle of NOAA/NCEI.