CENTRE FOR MARINE APPLIED RESEARCH

COASTAL MONITORING PROGRAM DATA MANAGEMENT PLAN

Prepared By
Nicole Torrie
Danielle Dempsey
Rachel Woodside
Leah Lewis-McCrea

Date
January 2025

Version
V2

Contact
P: 1-902-442-4660
E: info@cmar.ca
W: www.cmar.ca

A Data Management Plan created using DMP Assistant

Creators: Nicole Torrie, Danielle Dempsey, Rachel Woodside

Data Managers: Nicole Torrie, Danielle Dempsey, Rachel Woodside

Project Administrator: Leah Lewis-McCrea

Affiliation: Centre for Marine Applied Research (CMAR)

Funder: Nova Scotia Department of Fisheries and Aquaculture, Centre for Marine Applied Research

Template: Portage Template for Water Quality Research

Project start date: 2015-07-28

Last modified: 2025-01-10

1 Project Abstract

The Centre for Marine Applied Research (CMAR) operates a Coastal Monitoring Program (CMP) to fill a critical gap in ocean monitoring in Atlantic Canada. Through the CMP, CMAR collects data on ocean variables at monitoring stations within 1km of the coastline, and occasionally in rivers and lakes. The resulting datasets support science-based development of coastal industry, guide government policy and management decisions, encourage environmental stewardship, and help track coastal ecosystem changes. High-resolution data for the Coastal Monitoring Program is collected through three main program branches; Water Quality, Current, and Wave.

Through the Water Quality branch of the program, CMAR collects temperature, dissolved oxygen, and intermittent salinity measurements. These data are typically recorded by sensors deployed on vertical moorings referred to as "sensor strings". A typical sensor string configuration consists of a rope anchored to the seafloor and suspended in the water column by a sub-surface buoy (Figure 3). Sensors are secured to the rope at standard depths, typically 2, 5, 10, 15, etc., meters below the surface. Sensors are programmed to record every 1-60 minutes and are deployed for 3-12 months at a time. CMAR currently maintains sensors at roughly 70 locations and is continuously expanding the program into new areas throughout coastal Nova Scotia. Water Quality data has been collected at a total of 147 stations since the onset of the program (Figure 1).

Through the Current branch of the program, CMAR processes water speed and direction data collected using Acoustic Doppler Current Profilers (ADCPs) which are owned and deployed by the Nova Scotia Department of Fisheries and Aquaculture (NSDFA). Many of the ADCPs used for the Current branch of the program also collect wave height, wave period, and wave direction metrics for the Wave branch of the program. Each ADCP is mounted on the sea floor and is deployed for 1 – 3 months at a time. Current measurements are typically recorded every 10 minutes, and wave measurements every 1 hour. Current data have been collected at a total of 76 stations since the onset of the program, with wave data also collected at 58 of these locations.

In total, over 50,000,000 data observations have been collected between the three program branches since 2015. CMAR follows strict quality assurance (QA) and quality control (QC) procedures to ensure production of high-quality datasets. This Data Management Plan (DMP) is one quality assurance tool used by CMAR's Data Management Committee to ensure careful consideration for the entire CMP data lifecycle.

Relevant Links:
To learn more about the Coastal Monitoring Program or access the downloadable datasets, visit the Coastal Monitoring Program page on the CMAR Website.

To learn more about CMAR's data management practices, visit CMAR's Data Governance Website.

Table of Contents

2 Data Collection

2.1 Why are you collecting or generating your data?

The Coastal Monitoring Program (CMP) is an ongoing data collection effort in Nova Scotia that aims to support coastal industry, guide government decisions, encourage environmental stewardship, and ensure preparedness for climate change. The program was initiated by the Nova Scotia Department of Fisheries and Aquaculture (NSDFA) in the mid-2010s to assist with aquaculture site selection and management. At this time, resulting data were only processed and shared upon request, and the program lacked consistent data management, processing, and publishing procedures. CMAR assumed full responsibility for the Water Quality (temperature, dissolved oxygen, salinity) branch of the Program in 2019, and collaborates closely with NSDFA to process and publish data from the Current and Wave branches. CMAR has expanded the scope and mandate of the program beyond aquaculture, to support all coastal stakeholders. CMAR has published data collected through the Coastal Monitoring Program dating back to 2015. Data are now made publicly available in several formats and locations.

2.2 What types of data will you collect, create, link to, acquire and/or record?

Data collected is tabular observational data (various raw data formats depending on sensor types, processed into .csv formatted datasets), and photographs. Summary reports of the data from each branch are available on the CMAR website in .pdf format (Centre for Marine Applied Research, 2024c).

2.3 Where are you collecting or generating your data (i.e., study area)? Include as appropriate the spatial boundaries, water source type and watershed name.

The study area for the Coastal Monitoring Program includes coastal waters surrounding the province of Nova Scotia, Canada. Monitoring typically takes place within 1 km of the coastline, at depths from 0 to 80 m (Figure 1). Occasional monitoring may take place in Nova Scotia's lakes and rivers.

CMAR is also supporting efforts by other Canadian provinces (e.g. New Brunswick) to collect their own coastal data with the same field collection methods used for CMAR's Coastal Monitoring Program. Resulting data will be managed alongside the Nova Scotia data, following the same protocols outlined in this Data Management Plan.

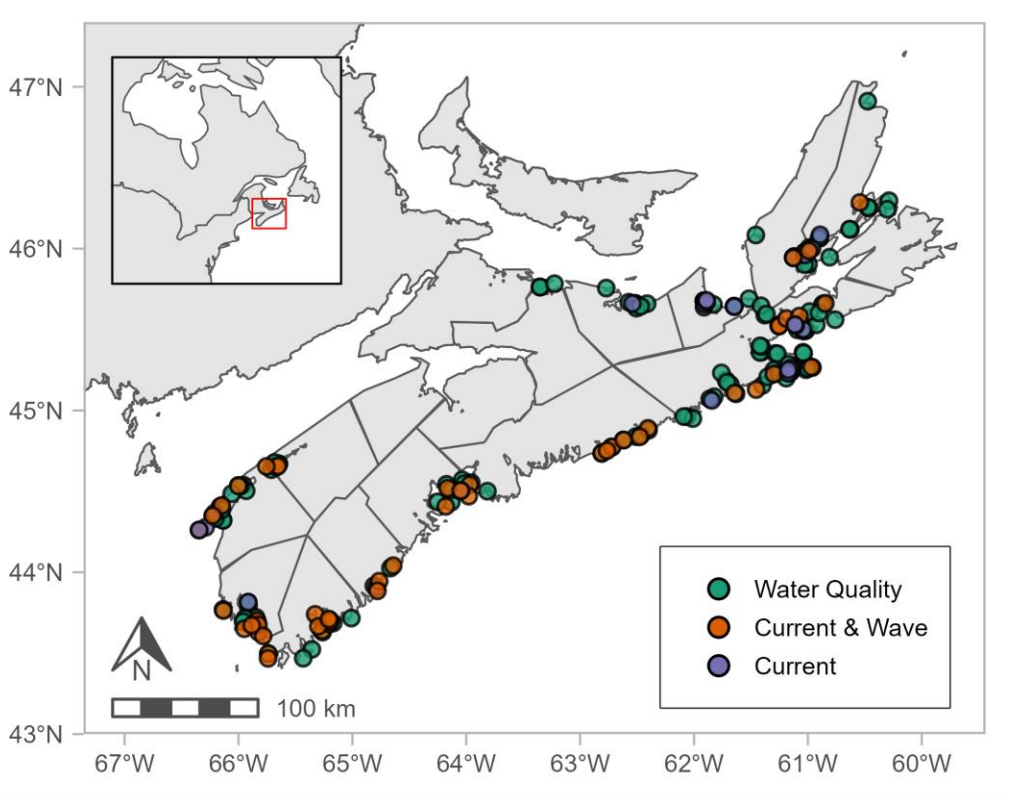

Figure 1 Coastal Monitoring Program study area and station locations

2.4 Are you using third party data? If so, describe the source of the data including the owner, database or repository DOIs or accession numbers.

Third party data collected by the Nova Scotia Department of Fisheries and Aquaculture (NSDFA) includes all Water Quality data collected prior to 2019, and all Current and Wave data. All datasets were provided to CMAR upon acquisition of the program in 2019. NSDFA continues to provide new Current and Wave data to CMAR, and CMAR is responsible for processing and publishing these data. All datasets with data pertaining to Nova Scotia's waters are approved by NSDFA before being published to the Nova Scotia Open Data Portal and CMAR website (Centre for Marine Applied Research, 2024c; Nova Scotia Open Data Portal, 2024).

Data collected by the New Brunswick Department of Agriculture, Aquaculture and Fisheries (NBDAAF) are third party data. These data are provided to CMAR after their collection. CMAR is responsible for data processing and publishing. CMAR is currently working to publish datasets to the New Brunswick Open Data Portal and CMAR website (Centre for Marine Applied Research, 2024b). All datasets with data pertaining to New Brunswick's waters will be approved by NBDAAF before being published.

CMAR may collect additional third-party data with other partners in the future.

3 Ethics and Legal Compliance

3.1 Does your project include sensitive data?

The Coastal Monitoring Program includes environmental data collected in partnership with industry stakeholders, including First Nations partnerships.

3.2 If your project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

Unprocessed (raw) data and associated metadata are securely stored on CMAR's server. This server can only be accessed by approved CMAR staff via a virtual private network (VPN). Processed data and associated metadata are stored in CMARs SQL database which is hosted on the server and only accessible to approved CMAR staff (password protected).

Before participation in the project, all partners are made aware that processed data will eventually be made publicly accessible. Some partners may not want their data distributed until a later date due to competition or conservation reasons; however, all agreements made with partners include an understanding that the data will be published after an agreed upon period of time. The exception to this is data collected in partnership with First Nations. First Nations data will not be made publicly available unless written consent is provided.

3.3 How will you manage other legal, ethical, and intellectual property issues?

All published Coastal Monitoring Program data are licensed under the Nova Scotia Open Government License.

4 File Management

4.1 What file formats will your data be in? Will these formats allow for data re-use, sharing and long-term access to the data? If not, how will you convert these into interoperable formats?

Data collected through the CMP undergo three phases with different storage types: 1. Raw data offloaded from the instruments; 2. Processed data stored in the CMAR database; 3. Datasets published on the Nova Scotia Open Data Portal (some of which are also published on CIOOS).

4.1.1 Raw Data

Raw data files are offloaded directly from the instrument, typically in a proprietary file format (Table 2). These are saved in the appropriate deployment folders on the CMAR server (see next section). Sensor-specific software is used to extract .csv (for sensor strings) or .txt files (for ADCPs) from the proprietary file format prior to data processing. CMAR data processing code is currently designed for .csv and .txt file formats; however, future iterations of the code may extract the data directly from the proprietary files. This would eliminate the interim step of extracting the flat files from the proprietary formats.

4.1.2 Processed Data

The raw data is compiled into a consistent format using CMAR R code (Table 1) and saved in the CMAR database (Figure 2). Quality control flags are applied to Water Quality data, and additional quality control is applied to the Current and Wave data to supplement the ADCP software checks. These quality control flags are stored with the data within the database. The CMAR database is a relational database built and administered using the free and open source software PostgreSQL and pgAdmin4 (pgAdmin, 2024; The PostgreSQL Global Development Group, 2024). It is hosted on a dedicated server and read and write access is controlled by the database administrator. The database can be queried by CMAR staff to review deployment information, assess inventory, and analyze measurements. For analysis of the data measurements, users should identify which quality flags to keep; otherwise, limited post-processing is required. This data can be used and re-used for the lifespan of CMAR.

Table 1. CMAR-developed R packages and their purpose

Package	Purpose
sensorstrings	Compile, format, and visualize CMAR Water Quality data.
qaqcmar	Apply Quality Control flags to CMAR Water Quality data.
calval	Validate calibrated sensors used in the Water Quality branch of the program.
adcp	Compile, format, visualize, and preliminary Quality Control for CMAR Current data
waves	Compile, format, visualize, and preliminary Quality Control for CMAR Wave data

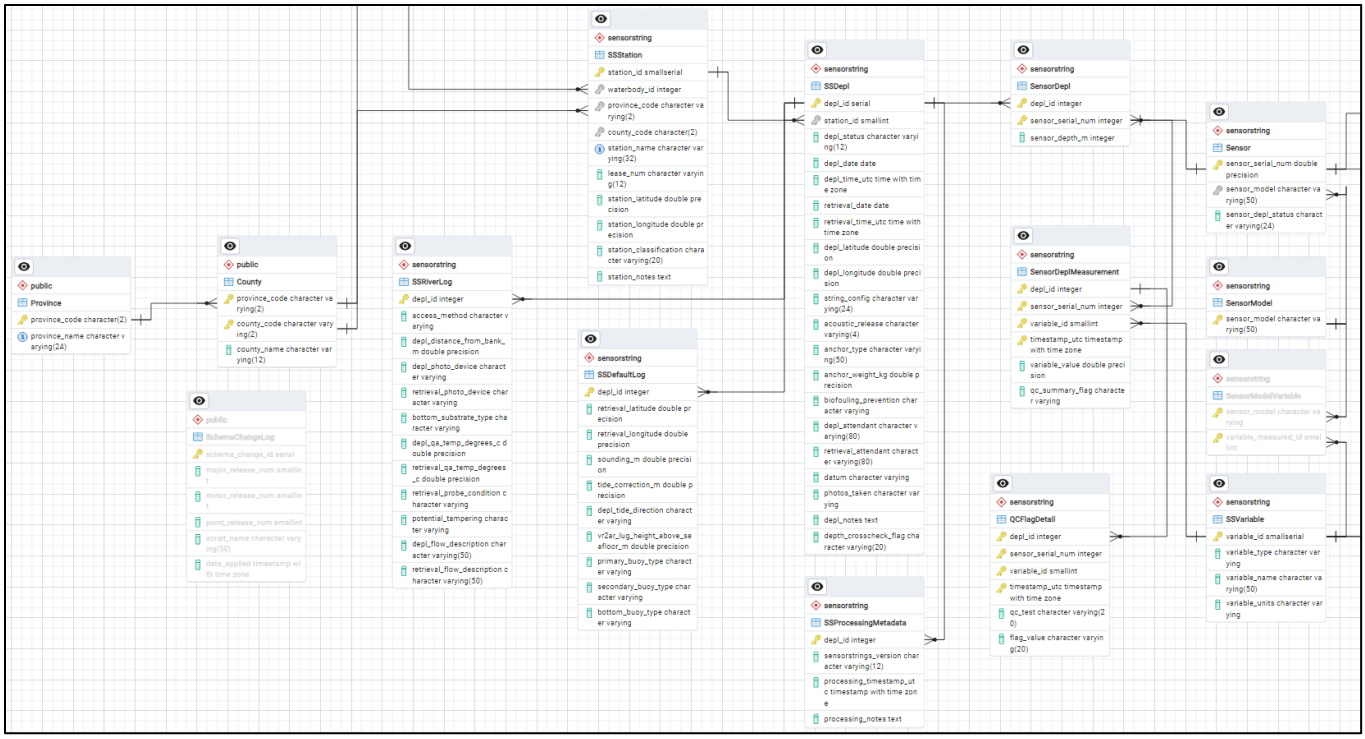

Figure 2. CMAR relational SQL database schema for Water Quality data

4.1.3 Published Datasets

To provide free access to the public, datasets are published annually. Water Quality, Current, and Wave datasets are grouped by county, with an associated "Station Locations" (Water Quality) or "Deployment Information" (Current and Wave) dataset. Datasets are assembled by querying the database and saving the results as a .csv file. All .csv files are sent to the Nova Scotia Open Data Portal for publication. Water Quality datasets are imported to the Canadian Integrated Ocean Observing (CIOOS) through the Nova Scotia Open Data Portal application programming interface. Datasets will be available for use and re-use for the lifespan of the platforms. Several download formats are available on each of these platforms, but it is recommended that users download the data in .csv format to ensure proper formatting and interoperability (Table 2).

Table 2. Data formats by sensor type

Sensor Type	**Program Branch**	**Data Extraction Software**	**Raw Data Format**	**Recommended Data Download Format**
Hobo Temp U22-001	Water Quality	HOBOware	.hobo	.csv
Hobo DO U26-001	Water Quality	HOBOware	.hobo	.csv
aquaMeasure DOT	Water Quality	aquaMeasure app	.csv	.csv
aquaMeasure SAL	Water Quality	aquaMeasure app	.csv	.csv
Vemco VR2AR	Water Quality	VUE	.vrl	.csv

Vemco VR2AR-X	Water Quality	VUE	.vrl	.csv
ADCP - Workhorse Sentinel 600kHz	Current and Wave	WavesMon and Velocity	.txt	.csv
ADCP - Sentinel V100	Current and Wave	WavesMon and Velocity	.txt	.csv
ADCP - Sentinel V50	Current and Wave	WavesMon and Velocity	.txt	.csv
ADCP - Sentinel V20	Current and Wave	WavesMon and Velocity	.txt	.csv

4.2 What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

Coastal Monitoring Program data are stored separately by program branch (i.e. Water Quality, Wave, Current). All files and folders are named in lowercase, with each space replaced by an underscore (i.e. snake case) for consistency. For each branch, data folders are organized by station name (e.g., meal_rock). Within each station folder, data for each deployment at that station are separated into a folder named with the station name and deployment date (e.g., meal_rock_2021-06-07). This deployment folder holds all raw data, data processing R scripts, and other supporting information such as fieldwork pictures. Raw data files are never altered directly. Instead, the R scripts used to process the data for each deployment are saved within the deployment folder so data can quickly be re-processed if needed (e.g., compile_meal_rock_2021-06-07.R), and all processing procedures are traceable. Any time metadata or R scripts are updated, old versions are moved to a folder called 'archive' within the deployment folder.

5 Documentation and Metadata

5.1 What documentation will be needed for the data to be read and interpreted correctly in the future?

5.1.1 Water Quality

To accurately interpret the Water Quality datasets, users need to know the deployment location details (station name, waterbody, latitude/longitude), sensor depths, sensor string configuration (see Figure 3), variable measured, units of measurement, the quality control flag value(s), and timestamp for each data observation. This information is contained within the columns associated with each observation, and column names are descriptive to reduce misinterpretation. A data dictionary is provided with each dataset on the Open Data Portal and can be downloaded in .xlsx format. This data dictionary contains descriptions, data type, units, and formatting examples for each column in the dataset. It is also important for users to note that the `sensor_depth_at_low_tide_m` value is approximate and is typically representative of sensor depth below surface at low tide, but may be interpreted differently if the sensors are attached to a floating configuration (Figure 3). Usage considerations are listed with each dataset on the Nova Scotia Open Data Portal and CIOOS Atlantic Platform. For example, click here to view the Guysborough County Water Quality Dataset usage considerations.

5.1.2 Current

To accurately interpret Current datasets, users need to know the deployment location details (deployment ID, waterbody, and latitude/longitude), ADCP configuration, variable measured, units of measurement, and timestamp for each observation. A description of each column, including associated units and data type, are available with each dataset on the Nova Scotia Open Data Portal. Each dataset is linked to an additional Deployment Information dataset which contains details of the ADCP deployment configuration. Usage considerations are listed with each dataset on the Nova Scotia Open Data Portal. For example, click here to view the Guysborough County Current Dataset usage considerations.

5.1.3 Wave

To accurately interpret Wave datasets, users need to know the deployment location details (deployment ID, waterbody, and latitude/longitude), sensor or buoy type and model, variable measured, units of measurement, and timestamp for each observation. A description of each column, including associated units and data type, are available with each dataset on the Nova Scotia Open Data Portal. Usage considerations are listed with each dataset on the Nova Scotia Open Data Portal. For example, click here to view the Guysborough County Wave Dataset usage considerations.

5.2 How will you describe samples collected?

5.2.1 Water Quality

CMAR uses several types of autonomous sensors to collect data for the Water Quality branch of the program, including temperature (degrees Celsius), depth (m), dissolved oxygen (% saturation or mg/L), salinity (practical salinity units), and/or acoustic detections from tagged animals (Table 3) (Centre for Marine Applied Research, 2023). CMAR typically programs sensors to record every 10 – 60 minutes to optimize battery life and memory storage, although other intervals may be used depending on the sensor and objective of the deployment. All sensors are retrieved to offload data (D. P. Dempsey et al., 2024).

Sensors are deployed on vertical moorings referred to as "sensor strings". A typical sensor string configuration consists of a rope attached to the seafloor by an anchor (e.g., scrap metal) and suspended in the water column by a sub-surface buoy. Sensors are secured to the rope at standard depths, usually 2, 5, 10, 15, etc. meters below the surface based on chart depth. Alternatively, sensors may be attached to other structures, including floating docks, buoys, and other infrastructure (e.g., aquaculture cages) (Figure 3, Table 4). Sensor strings are deployed at a station for several months to a year, depending on battery life, memory storage, and maintenance requirements.

The purpose of the deployment and characteristics of the station location dictate the string design, including anchor weight, buoyancy, number and type of sensors, and configuration. Sensor strings are generally deployed 200 m to 1000 m from shore, in depths up to 80 m. Historically, strings were deployed with only temperature sensors, but since 2019 a dissolved oxygen sensor is often attached 5 m below the surface. At the request of industry, salinity sensors are also included at stations near proposed and existing shellfish aquaculture leases.

Once retrieved, data from each sensor on a sensor string are processed together and uniquely identified by the deployment location (station name) and deployment start date (e.g. Long Beach 2024-08-22). Subsequent deployments may occur in the same location, and the deployment start date is used to differentiate data from separate deployments. For each deployment, metadata includes the deployment location (station name, waterbody, and latitude/longitude), sensor depths, sensor models, configuration, deployment date, retrieval date, deployment sounding, retrieval method, and other notes.

Table 3. Water Quality sensor details

Sensor Type	Program Branch	CMAR Measurement Interval	Manufacturer	Variable(s) Measured
Hobo Temp U22-001	Water Quality	15 minutes	Onset	Temperature
Hobo DO U26-001	Water Quality	15 minutes	Onset	Temperature, Dissolved Oxygen
aquaMeasure DOT	Water Quality	10 minutes	Innovasea	Temperature, Dissolved Oxygen, Sensor Depth
aquaMeasure SAL	Water Quality	10 minutes	Innovasea	Temperature, Salinity
Vemco VR2AR	Water Quality	1 hour	Innovasea	Temperature, Acoustic Detections, Sensor Depth
Vemco VR2AR-X	Water Quality	1 hour	Innovasea	Temperature, Acoustic Detections, Sensor Depth

Table 4. Water Quality deployment configuration descriptions

Configuration	Description	Retrieval Method
Sub-surface buoy	Attached to sub-surface buoy. Typically has an acoustic release, but sometimes retrieved by diver or drag line. Does not float with the tide.	Trigger acoustic release.
Surface buoy	Attached to a surface buoy. Floats with the tide.	Remove with winch.
Attached to gear	Typically attached to aquaculture gear, but may be attached to another structure. Floats with the tide.	Cut from equipment.
Attached to fixed structure	Typically attached to a wharf, but may be attached to a bridge or pole. Does not float with the tide.	Cut from structure.
Floating dock	Attached to a floating dock. Floats with the tide.	Cut from dock.
Unknown	Configuration not indicated in historical metadata.	

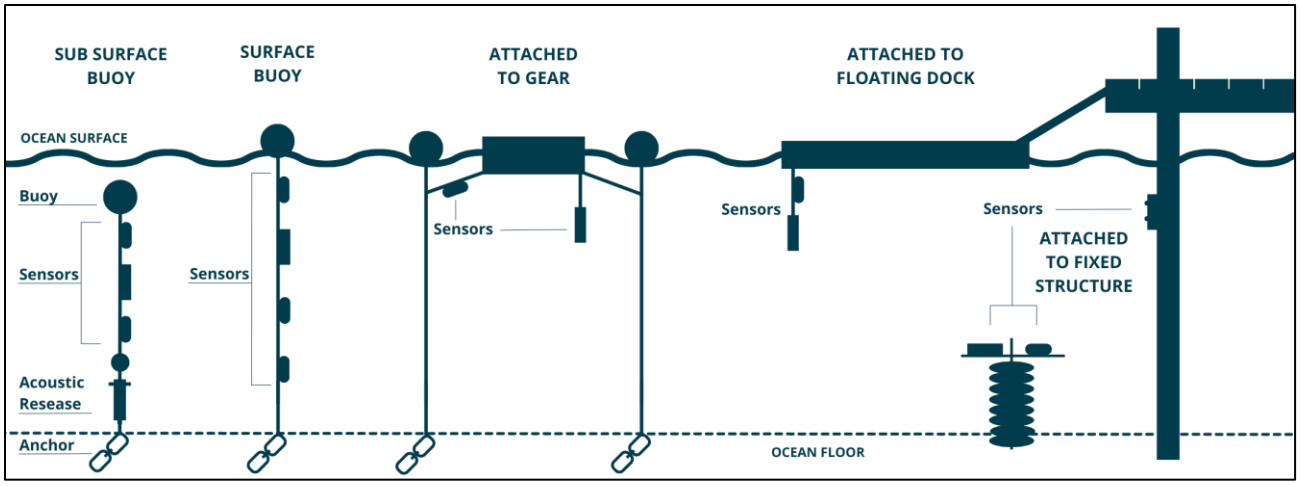

Figure 3 Sensor deployment configurations for the Water Quality branch (not to scale)

5.2.2 Current & Wave

NSDFA maintains responsibility for most of the Current and Wave workflow, including site selection (with input from CMAR), deployment, retrieval, and data offload. Current and Wave parameters are monitored using Acoustic Doppler Current Profilers (ADCP), which are instruments that use sound to measure water velocities over a range of depths (Teledyne RD Instruments, 2017). ADCP manufacturer software averages raw observations to calculate current speed and direction over a given time interval. Some units can be configured to measure sea state parameters including wave height, period, and direction (RD Instruments, 2017). NSDFA uses several ADCP instrument models, including the Sentinel V20, Sentinel V50, Sentinel V100, and Workhorse Sentinel 600kHz. Each ADCP is mounted on the seafloor and is deployed for 1 – 3 months (Figure 4; (Baron, 2020)). Current measurements are typically recorded every 10 minutes, and wave measurements every 1 hour. Instruments must be retrieved to offload the data. The instruments are retrieved by activating an acoustic release or with drag lines. NSDFA processes the raw data with Velocity and WavesMon4 Software by Teledyne RD Instruments (Teledyne RD Instruments, 2017), and sends the processed data to CMAR.

Deployment data are uniquely identified by the deployment location (station name) and deployment start date (e.g. Shut-In Island 2020-07-03). Subsequent deployments may occur in the same location and the deployment start date is used to differentiate data from separate deployments. A unique deployment ID is also assigned to each deployment. For each deployment, metadata includes deployment location (station name, waterbody, and latitude/longitude), sensor model, deployment date, retrieval date, deployment sounding, bin depth, and other notes.

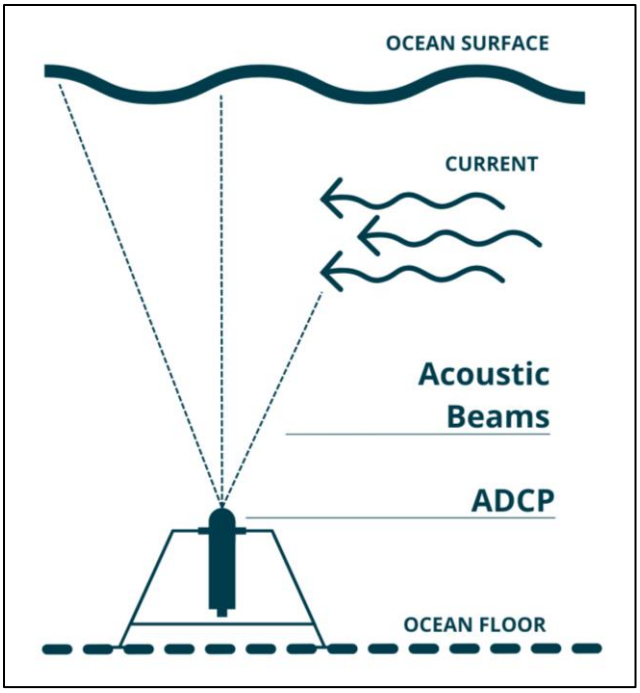

Figure 4 ADCP deployment configuration for the Current and Waves branches (not to scale)

5.3 How will you analyze and interpret the water quality data?

Data are available to CMAR for internal use and analysis as needed. Internally, data can be accessed through the CMAR database and analyzed within R. CMAR plans to incorporate and analyze the data as new projects and funding opportunities arise; however, the main program priority is to continue to produce high quality, public datasets for others to perform their own analyses with.

Coastal Monitoring Program data are publicly available in several formats for other ocean users to analyze and interpret. Water Quality datasets are available for download in .csv (recommended) format from the Nova Scotia Open Data Portal and CIOOS Atlantic. Current and Wave datasets are available for download in .csv (recommended) format from the Nova Scotia Open Data Portal. Once downloaded, datasets can be analyzed with any software of choice (e.g. R, Python, Excel).

5.4 What kind of Quality Assurance/Quality Control procedures are you planning to do?

As part of CMAR's data management initiatives, CMAR ensures careful Quality Assurance (QA) and Quality Control (QC) practices are followed and well documented. Quality Assurance includes processes that are employed to support the generation of high-quality data (e.g., sensor calibration and validation, upkeep of Standard Operating Procedures (SOPs)), while Quality Control describes the steps that support the delivery of high-quality data (e.g., automated data

flags, manual review) (Bushnell et al., 2019; IOOS, 2020b). CMAR maintains nearly 40 detailed SOPs to document all procedures, including deployment planning, fieldwork, and data processing.

5.4.1 Water Quality

Prior to sensor deployment, CMAR calibrates all sensors that can be calibrated, and then completes a series of validation tests on all sensors. These tests ensure each sensor is recording within an expected range for each variable (Torrie, Dempsey, & Woodside, 2024). CMAR has developed the R package 'calval' to evaluate and visualize sensor performance during each validation test (Torrie et al., 2024). Only sensors that pass all validation tests are cleared for deployment (Torrie et al., 2024).

After sensors are retrieved from deployment, CMAR applies several quality control tests during data processing. These tests and data flagging scheme were adapted from recommendations by the US Integrated Ocean Observing System (IOOS, 2018, 2020a, 2020b). Five automated tests and an additional "human in the loop" test are applied to all the Water Quality data using the 'qaqcmar' R package (Dempsey, Torrie, & Woodside, 2024). For each data observation, the worst flag from all tests is published with the dataset, and users can filter the data set for records that meet their quality criteria (UNESCO, 2013).

Extensive documentation of all QA/QC procedures applied to the Coastal Monitoring Program Water Quality datasets, including quality control test details, can be found on the CMAR Data Governance Website (Centre for Marine Applied Research, 2024b).

5.4.2 Current and Wave

NSDFA is responsible for ADCP pre-deployment, fieldwork, and data offload procedures. The ADCP software performs many quality control checks prior to data offload. Some additional quality control checks are applied during CMAR data processing. Efforts are ongoing to develop and apply more formal tests and to publish the resulting flags with the Current and Wave datasets.

5.5 How will you make sure that documentation is created or captured consistently throughout your project?

CMAR maintains nearly 40 detailed SOPs to document all procedures, including deployment planning, fieldwork, and data processing. CMAR SOPs are stored on the CMAR server. These SOPs are reviewed by staff annually and updated as needed. All CMP staff are required to familiarize themselves with and follow relevant SOPs. In addition, manuals for each sensor are stored on the CMAR Google drive for easy access and review. See the question "How will you make sure that metadata is created or captured consistently throughout your project?" below for additional information about metadata documentation procedures.

The CMP Data Management Committee meets weekly to discuss data management initiatives and ensure everyone is up to date on data management decisions and procedures. The full CMP team meets monthly to share progress, updates, and pain points. Any major changes to documentation procedures are discussed during these meetings.

5.6 List any metadata standard(s) and/or tools you will use to document and describe your data:

CMAR's Coastal Monitoring Program metadata files provide information required by Nova Scotia Open Data Portal and CIOOS Atlantic.

Stations and waterbodies are named in accordance with the Canadian Geographical Names Database.

Dataset columns are named in accordance with the Climate and Forecast standard naming conventions.

Datetime data are published in accordance with the ISO 8601 datetime format, and are always in UTC.

5.7 How will you make sure that metadata is created or captured consistently throughout your project?

CMAR uses a deployment field sheet template to capture essential metadata components for all sensor deployments. Similarly, a consistent paper retrieval field sheet template is used for all sensor retrievals. If sensors are in a location where they are easily accessible for cleaning (i.e. on an aquaculture lease), staff who will be conducting the cleaning procedures are provided with a sensor maintenance log field sheet to keep track of cleaning. Field sheets are scanned quarterly, and the resulting images are saved on the CMAR server.

Metadata from the field sheets are digitized into a metadata tracking sheet in .xlsx format, from where they are imported into CMAR's relational SQL database. The metadata tracking sheet has conditional formatting, autofill, and dropdowns where possible to ensure consistent data formats and naming conventions are used. These reduce the number of typos and misspellings in the metadata. The CMAR database also requires specific data formats for many of the data and metadata columns, and will flag issues in the data or metadata before they are imported into the database. Any subsequent changes made to the metadata are made within the database, and previous versions are stored in history tables within the database.

6 Storage and Backup

6.1 How and where will your data be stored and backed up during your research project?

Data and associated metadata files, including the database, are stored on the CMAR server. This server is accessible by approved CMAR staff via a virtual private network (VPN). The server is backed up each weekday by Perennia IT in multiple locations. This includes backup to local storage devices in three Perennia office locations, and a remote backup through cloud storage provider Acronis.

6.2 What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

Currently, roughly 463 gigabytes of storage space have been filled on the CMAR server. The Coastal Monitoring Program data from all program branches requires roughly 70 gigabytes of this storage space. Other items stored on the server include metadata, photographs, presentations, tracking sheets and additional documents related to the program. It is anticipated that this program will continue to expand, and data will be stored for at least 5 years.

6.3 How will the research team and other collaborators access, modify and contribute data throughout the project? How will data be shared?

Water Quality data are uploaded to the CMAR server by CMAR's Field Technician after equipment is retrieved from the field. Occasionally, other collaborators (e.g. researchers, aquaculture lease owners) deploy sensors for CMAR. These sensors are returned to CMAR after the deployment period is over, and data are offloaded and uploaded to the CMAR server by CMAR's Field Technician. This server is accessible only via Perennia's VPN, and can be accessed and altered by approved CMAR staff. Current and Wave data are shared with CMAR through secure file transfer or email depending on the file size and then saved to the appropriate folder on the server.

On the CMAR server, data are sorted based on which program branch it was collected under. Once on the CMAR server, raw data files are not altered directly. All data processing takes place in R, and all outputs and associated R scripts are saved separately. Processed data are stored within CMAR's SQL database. CMAR staff primarily interact with the database using R code to query and work with water quality data. If unpublished data files need to be shared with anyone outside of CMAR, they are shared via secure file transfer, or through email if the file sizes are small enough. Published datasets are available online for free in several formats.

7 Responsibilities and Resources

7.1 Who will be responsible for managing this project's data during and after the project, and for what major data management tasks will they be responsible?

The primary responsibilities of data management fall under responsibilities of the Field Technician(s), Principal Scientist, Coordinator, and Database Developer. Descriptions of all roles related to the Coastal Monitoring Program are described below:

7.1.1 Project Administrator

The Project Administrator role is filled by CMAR's Research Director.
Responsibilities of the Project Administrator include:

1. Obtain program funding
2. Manage program budget and expenses
3. Hire and manage staff (Data Managers, Field Technician)
4. Collaborate with stakeholders over monitoring priorities (variables and locations)
5. Guide program priorities and scheduling as needed

7.1.2 Contracted Field Team

CMAR uses a contractor model for the Field Team. A minimum hired crew of two is required to operate the boat for each field trip.
Responsibilities of the Field Team include:

1. Plan and mobilize for fieldwork (check weather, ramps, accessibility, etc.)
2. Navigate to sensor stations, assist with gear deployment/retrieval
3. Maintain boat and related equipment

7.1.3 Field Technician and Research Assistant

The Field Technician and Research Assistant role is filled by one member of the CMAR staff. Occasional tasks may be passed off to other CMAR staff or interns depending on the volume of work to be done at any given time.
Responsibilities of the Field Technician and Research Assistant include:

1. Conduct pre-deployment sensor calibration
2. Conduct pre-deployment sensor validation tests
3. Communication with contracted field team
4. Prepare equipment for deployment
5. Fieldwork (deployment and retrieval of equipment)
6. Collect detailed and accurate field metadata
7. Conduct post-deployment sensor validation tests

8. Offload deployment data
9. Upload deployment data, supporting files and metadata to the CMAR server
10. Maintain sensors and equipment
11. Keep up-to-date sensor inventory
12. Create and update related SOPs as needed
13. Test and recommend improvements to methods and operating procedures

7.1.4 Data Manager(s)

The Data Manager role at CMAR can be filled by multiple people, depending on the volume of work to be done at any given time. A minimum of 2 data managers are recommended to maintain the program at its current scale. Currently the Data Manager role is split into the following positions: Coastal Monitoring Program Coordinator, Coastal Monitoring Program Principal Scientist, and Database Developer. Occasional tasks may be passed off to other CMAR staff or interns depending on the volume of work to be done at any given time.

Responsibilities of the Data Manager(s) include:

Shared Responsibilities

1. Participate in Data Management Committee
2. Process deployment data
3. Attend conferences, presentations, tradeshows and events to communicate program data

Coastal Monitoring Program Coordinator

1. Write and oversee tender for Field Team, and resulting contract
2. Create and oversee fieldwork schedule
3. Approve and manage gear orders
4. Create and maintain the Data Management Plan (DMP)
5. Create and update SOPs as needed
6. Develop and maintain R code for sensor validation
7. Evaluate validation data
8. Troubleshoot and resolve field-related issues (e.g. gear loss or entanglement)

Coastal Monitoring Program Principal Scientist

1. Keep track of data processing status and timelines
2. Develop and maintain R code for data processing, quality control, and analysis
3. Work towards improving data quality (improve methods, quality control code, etc.)
4. Assemble and publish datasets
5. Generate and maintain summary reports for all program branches
6. Perform data analysis for various internal initiatives
7. Lead Data Management Committee
8. Supervise summer intern

Database Developer

1. Develop and maintain CMAR database
2. Train staff to upload and access data within the CMAR database

7.2 How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

In the event of staff turnover, responsibilities will be transitioned to new and/or existing staff. Typically, at least two staff are familiar with each procedure to ensure knowledge is not lost. Responsibilities are also extensively documented in CMAR's SOPs. The project administrator will be primarily responsible for transferring these SOPs to the new data manager.

7.3 What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

CMAR owns a sensor inventory valued at over $800,000 for the Water Quality branch of the program. On top of the initial purchase cost of this equipment, there are ongoing annual costs related to fieldwork, storage, new equipment purchases, and equipment maintenance adding up to roughly $250,000/year. Equipment and fieldwork costs for the Current and Wave branches of the program are covered directly by NSDFA. Salaries for CMAR staff working full time on this program (including full time Field Technician and Research Assistant, full time Principal Scientist, full time Coordinator, contracted part-time Database Developer, and a part-time Data Technician Intern) amount to roughly $260,000/year. The SQL server which hosts the CMAR database is rented at a rate of roughly $1,300/year (including hardware, backup, and storage). This brings total annual operating costs for the program to just over $510,000, with data-management-related costs being $261,300 of this.

8 Sharing, Reuse and Preservation

8.1 Do you, your institution or collaborators have an existing data sharing strategy?

In accordance with the FAIR data principles, all processed data collected for the Coastal Monitoring Program are made publicly available for download via the Nova Scotia Open Data Portal and processed Water Quality data are available on the CIOOS Atlantic platform (Wilkinson et al., 2016). Data are uploaded annually on these platforms. Summary reports of the data are also uploaded to the CMAR website and updated annually.

8.2 Are there restrictions on sharing due to ethics or legal constraints?

There are no restrictions on sharing the Coastal Monitoring Program data, unless data has been collected in partnership with First Nations. First Nations data will only be shared if written consent is given.

8.3 What data will you be sharing and in what form? (e.g., raw, processed, analyzed, final).

Data are shared in processed form. In this format, data from all sensors deployed in each county has been combined into one dataset, and any outliers have been flagged. For each Water Quality data observation, the highest (i.e. worst) flag from all QC tests is published with the dataset, and users can then filter the dataset for records that meet their quality criteria. Extensive documentation of all QA/QC procedures applied to the Coastal Monitoring Program Water Quality datasets, including quality control test details, can be found on the CMAR Data Governance Website (Centre for Marine Applied Research, 2024b).

8.4 Will you deposit your data for long-term preservation and access at the end of your research project? Please indicate any repositories that you will use.

Water Quality data are published to the Nova Scotia Open Data Portal and CIOOS Atlantic platforms. Current and Wave data are published to the Nova Scotia Open Data Portal.

8.5 What steps will you take to ensure your data is prepared for preservation?

Datasets are publicly available on the Nova Scotia Open Data Portal and CIOOS Atlantic (where applicable) platforms from where they can be directly downloaded in non-proprietary .csv file formats. Metadata and usage considerations have been extensively documented and are available with the datasets on these respective platforms. The "Data Access Reference Sheet" and Station

Locations Map on the CMAR website assist with access to the various data links (Centre for Marine Applied Research, 2024a, 2024c).

Methods have been documented in several locations including the Coastal Monitoring Program Data Governance website (Centre for Marine Applied Research, 2024b), and the paper "The makings of a high resolution Coastal Monitoring Program" (D. P. Dempsey et al., 2024).

8.6 What type of end-user license will you use for your data?

The Coastal Monitoring Program data are licensed under the Nova Scotia Open Government License.

8.7 What steps will be taken to help the research community know that your data exists?

The data will be made accessible through the Nova Scotia Open Data Portal and CIOOS Atlantic (where applicable) platforms. CMAR will continue to advertise the Coastal Monitoring Program at various events, conferences and presentations, as well as in social media posts when possible. CMAR has developed a paper "The makings of a high resolution Coastal Monitoring Program" to further describe the program background and methods (D. P. Dempsey et al., 2024). CMAR will also continue to release articles in local newsletters and magazines such as the SeaFarmers Magazine, and the COVE news board on their website.

9 References

Baron, A. (2020). *ADCP Deployment Length Guidelines*. Retrieved from https://cmar.ca/wp-content/uploads/sites/15/2021/07/2018-10-23-DSA-DFA-ADCPDeploymentLengthGuidelines_FINAL.pdf

Bushnell, M., Waldmann, C., Seitz, S., Buckley, E., Tamburri, M., Hermes, J., . . . Lara-Lopez, A. (2019). Quality Assurance of Oceanographic Observations: Standards and Guidance Adopted by an International Partnership. *Frontiers in Marine Science, 6*(706). doi:10.3389/fmars.2019.00706

Centre for Marine Applied Research. (2023). CMAR Coastal Monitoring Program Reference Sheet: Water Quality Sensor Details. In (3 ed.).

Centre for Marine Applied Research. (2024a). CMAR Coastal Monitoring Program Reference Sheet: Report and Data Access. In.

Centre for Marine Applied Research. (2024b). CMAR Data Governance Website. Retrieved from https://dempsey-cmar.github.io/cmp-data-governance/pages/cmp_about.html

Centre for Marine Applied Research. (2024c). Coastal Monitoring Program Station Locations Map. Retrieved from https://cmar.maps.arcgis.com/apps/webappviewer/index.html?id=fa287056727e42cf8e57e71b967bf563

Dempsey, D., Torrie, N., & Woodside, R. (2024). qaqcmar: Apply Quality Control Flags to CMAR Water Quality Data (Version 1.06). Retrieved from https://github.com/dempsey-CMAR/qaqcmar

Dempsey, D. P., Torrie, N. L., Reid, G. K., Lewis-McCrea, L. M., McKee, A., & Woodside, R. (2024, 23-26 Sept. 2024). *The Makings of a High Resolution Coastal Monitoring Program.* Paper presented at the OCEANS 2024 - Halifax.

IOOS. (2018). QARTOD Manual for Real-Time Quality Control of Dissolved Oxygen Observations. In (Vol. Version 2.1).

IOOS. (2020a). QARTOD Manual for Real-Time Oceanographic Data Quality Control Flags. In.

IOOS. (2020b). QARTOD Manual for Real-Time Quality Control of In-situ Temperature and Salinity Data: A Guide to Quality Control and Quality Assurance for In-situ Temperature and Salinity Observations. In.

Nova Scotia Open Data Portal. (2024). Coastal Monitoring Program Search Results. Retrieved from https://data.novascotia.ca/browse?tags=coastal+monitoring+program

pgAdmin. (2024). pgAdmin: PostgreSQL Tools. Retrieved from https://www.pgadmin.org/

RD Instruments. (2017). Waves Primer: Wave Measurements and the TRDI ADCP Waves Array Technique. In.

Teledyne RD Instruments. (2017). Velocity Software User's Guide. In.

The PostgreSQL Global Development Group. (2024). PostgreSQL: The World's Most Advanced Open Source Relational Database. Retrieved from https://www.postgresql.org/

Torrie, N., Dempsey, D., & Woodside, R. (2024). calval: Validate Calibrated Sensor Performance. Retrieved from https://github.com/ntorrie/calval/tree/main

UNESCO. (2013). Recommendation for a Quality Flag Scheme for the Exchange of Oceanographic and Marine Meteorological Data. In (Vol. 3).

Wilkinson, M., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*. doi:10.1038/sdata.2016.18