Ocean Alkalinity Enhancement Data Management Protocol

Revision History

Date	Version	Revision Description	Notes
01/23/2025	0.1.0	Original draft document published for public review	

The first open comment review period is January 23, 2025 - March 07, 2025.
Notes for reviewers are highlighted in orange font.

Acronyms and Abbreviations

mCDR	Marine carbon dioxide removal
OAE	Ocean alkalinity enhancement
DOI	Digital object identifier

Definitions of Selected Terms

Metadata	Metadata is structured information that describes and provides context for a data resource, helping to ensure that the dataset remains discoverable and usable in the future.
Data standards	Data standards are a set of agreed-upon rules, formats, and conventions used to define and structure data, ensuring consistency, interoperability, and clarity across different systems, datasets, and organizations. They help maintain data quality, facilitate data sharing, and enable effective analysis by establishing uniformity in how data are collected, represented, and exchanged.

10c	Column header names	Standardized column header terms describing a parameter, these may be an abbreviation of the measured parameter.
10d	Controlled vocabulary	Controlled vocabularies are standardized lists of terms and definitions used to ensure consistency in the naming and classification of concepts within a specific domain. By limiting the use of predefined terms, controlled vocabularies help avoid ambiguity, enhance data interoperability, and improve the accuracy of data retrieval and analysis across different systems and datasets.
10e	Ocean alkalinity enhancement	Ocean alkalinity enhancement (OAE) is a climate change mitigation strategy that involves increasing the alkalinity of seawater to enhance its capacity to absorb and store atmospheric carbon dioxide.
10f	Platform	Any physical structure or system used to support and deploy instruments, sensors, or other equipment for collecting data in the ocean. Platforms can include research vessels, ships of opportunity (SOOP), profiling floats, buoys, underwater vehicles, and moorings.
10g	Sensor data	Sensors refer to instruments or devices used to measure and collect data on various physical, chemical, and biological parameters of the ocean environment and are typically deployed on ships from rosettes or underway systems, buoys, underwater vehicles, or autonomous platforms. Data collected from these systems are considered sensor data and do not refer to data from autosampling devices.
10h	Quality control	Methods or procedures involving validating and verifying collected data to identify and correct errors, inconsistencies, or outliers, ensuring that the measurements are accurate and suitable for analysis. This typically includes tasks such as instrument calibration, data validation checks, and cross-referencing with other datasets to

		maintain the integrity of scientific results.
10i	Model data	The model data referenced in this protocol refers to code, configuration and output from mathematical simulations that discretise the equations for fluid motion and energy transfer and integrate these over time on a realistic three-dimensional grid. This encompasses model output relevant to OAE projects on nearfield and regional scales, as well as global circulation models (GCMs) and Earth System Models (ESMs). This could include ocean circulation models with or without coupling to biogeochemical, sediment, sea ice, or atmospheric models. This does not currently cover data standards for conceptual, process models, 1D or 2D models, or simplified plume mixing zone models.
10j	Data file	Refers to a file containing values of some measurements. File format type may vary (e.g., NetCDF, xlsx, xml), however all data files will contain quantitative values with associated column header names.
10k	Baseline	Baseline refers to the initial set of data or conditions that are recorded before any interventions or modifications are made to the ocean environment. This baseline field data serve as a reference point for comparing future measurements, allowing for the assessment of the effectiveness and impacts of the OAE interventions over time, such as changes in ocean alkalinity, CO_2 absorption, or ecosystem health.
10l	Intervention	An intervention refers to the intentional action or process applied to the ocean to alter its chemical or physical properties in order to enhance its capacity for carbon dioxide removal. This could include adding alkaline substances to the water or implementing other methods aimed at increasing ocean alkalinity and improving the ocean's ability to absorb and store atmospheric CO_2.
10m	Control	A control site refers to a designated area in proximity to an intervention site, with shared

	characteristic waters, but that remains unaffected by the OAE intervention, serving as a baseline for comparison during intervention. The purpose of a control site is to isolate and account for natural variability in oceanographic conditions, biogeochemical processes, and carbon fluxes, enabling the evaluation of changes directly attributable to OAE activities.

Additionally, observation types are provided in Jiang et al., 2023b and are archived by NOAA for reference (link).

Background and Introduction

This document outlines recommendations for producing consistent data and metadata for Ocean Alkalinity Enhancement (OAE) field trial projects. Its first iteration was produced in partnership with the ocean sciences community, with the intention to remain a **living document** and continually improve to reflect best known scientific practice.

Objectives & Guiding Principles

The objective of the OAE Data Management Guidelines and this OAE Data Protocols document are to enable marine Carbon Dioxide Removal (mCDR)-OAE data collected from academia, government, non-profit, and industry to be documented in a consistent way, and make them findable and discoverable from shared data repositories to facilitate future data synthesis efforts.

OAE introduces unique challenges and opportunities for data standardization. Traditional oceanographic data standards, while robust, require updates to address the specific needs of OAE projects.

The main updates and recommendations are driven by these Guiding Principles:

- **Project Comparability:** Developing guidelines that ensure data from different OAE projects are intercomparable, enabling meta-analyses and large-scale assessments of OAE effectiveness and environmental impact.
- **Minimal Burden on Data Providers:** Establishing streamlined protocols and tools that simplify data submission while ensuring high-quality, standardized outputs.
- **Flexibility for Innovation:** Allowing for innovation in project designs by creating standards that accommodate diverse methodologies and intervention scales without imposing restrictive requirements.
- **Transparency and Accessibility:** Promoting open and transparent data sharing, with appropriate metadata, to facilitate peer review, collaboration, and public trust in OAE efforts.

By building on existing standards and addressing these updates, the goal is to create a system that supports rigorous science while remaining practical and adaptable for data providers. This ensures that OAE projects contribute meaningfully to the collective understanding of marine carbon dioxide removal while fostering collaboration across the community.

Methodology

This protocol was developed by the OAE community in a multi-process method, starting with a workshop during OCEANS 2024 with participants representing academia, government, non-profit, and industry to gather initial feedback and input for sensor output, model output, and discrete carbonate and nutrient data. This feedback was developed into a first outlined draft of recommendations, which were reviewed by attendees and developed into the first draft OAE Data Protocols 0.1.0.

Working Groups were formed to capture input from biological sciences, sediment processes, and social science data, each hosting virtual meetings to gather community feedback, which further informed the initial draft. The recommended variable column header names and controlled vocabularies are made to mirror existing naming conventions, and long-standing recommendations by the Ocean Acidification Community. The draft was then provided to the Steering Team members for internal review. The resulting draft was presented at a second workshop during AGU in December 2024 for additional mCDR-community feedback, followed by additional internal revisions to create the draft that was presented during an open public review period. This document represents the final conclusions developed by the OAE community to ensure projects will be standardized, findable, openly accessible, and intercomparable.

We are grateful for the contributions of the workshop participants, working group members, the Steering Committee and those who provided comments during the open review period. For a list of contributors and contributing authors, please see the *Acknowledgements* section.

Intended users

This protocol is designed to assist data producers involved in OAE projects that include model output, and/or field data from commercially available sensors, discrete observations, social sciences, sediment processes, and/or biological processes, in consistently documenting their datasets, while also providing guidance on selecting appropriate repositories, submission timelines, controlled vocabularies, and best practices for ensuring data intercomparison.

Projects compliant with this protocol will at minimum meet the requirements below, however additional recommendations provided are strongly encouraged. The guidelines and recommendations outlined in this document may be applicable to a wider range of mCDR methods outside of OAE. The protocols outlined are presented as best practice recommendations, leaving the decision to adopt them at the discretion of individual mCDR researchers.

The protocol is organized into five chapters:

1. **Metadata** - Building on existing metadata guidelines that accompany each repository, this protocol requires *OAE Metadata* to provide essential context unique to OAE field trials. It also includes a *Model Description Template* to introduce and define key fields for documenting model and model output data details.
2. **Guidelines for Data Submission** - This chapter outlines how to set up, manage and submit data across a range of data types.
3. **Column Header Names and Controlled Vocabularies -** Recommendations are provided for column header names as well as controlled vocabularies and definitions, including OAE-specific fields.
4. **Emerging Standards -** This section outlines future work of the Data Management Guidelines project.
5. **Deprecated Standards and Practices -** Versioning control for the Data Protocol is outlined here.

Metadata

Depending on your choice of repository, there may be a structured metadata interface required with data submission. In cases where this is optional, it is recommended to refer to the metadata contents provided by the NOAA OCADS (Jiang et al., 2023a) metadata form (https://www.dropbox.com/scl/fi/8lrsyufqljs1qi7ut9fji/Metadata_template_4.1.xlsx?rlkey=nfusfv7cw3khsrmrvaup7zf88&st=2v3c69sf&dl=0). This form includes both metadata that are general to any oceanographic project (e.g., cruise_id, experiment_id, data_submitter, investigators, etc.), specific metadata for discrete, sensor, and model data, as well as OAE-specific fields that have been recommended by the OAE community and are presented below (see *OAE Metadata* for field data, and *Model Description Template* for model data). If provided as a separate file, this information can be provided in xlsx, netCDF, json, or XML format and can be named 'metadata'.

OAE Metadata

mCDR research, particularly OAE, involves creating deliberate perturbations to study their effects. Given the diversity of approaches to raising alkalinity or reducing acidity, along with the site-specific nature of OAE field data, additional metadata fields are needed on top of existing metadata templates that are designed for ocean carbon and acidification studies. OAE Metadata provides the necessary context for the research community to effectively find, interpret, compare, and use the data across different studies and methodologies.

For ease of use, the OAE-relevant fields are provided in the *OAE Metadata* table below. Note that most of these OAE-relevant metadata fields are integral to the updated OCADS metadata template except for modeling output studies. If you are submitting your data to a repository that has required structured metadata fields, the information below can be provided in plain text, netCDF, or XML format as a separate upload titled ‘OAE_metadata_supplement’. If there is no required metadata structure, please refer to the NOAA-OCADS full metadata table (see *Metadata*).

Note to reviewers: Our goal is to identify key fields and examples that make the metadata useful and ***searchable****, without overwhelming the data submitter. Please review for missing fields, missing [inputs] that you’d like to be searchable, inconsistencies in vocabulary and areas that are too detailed.* ***We are also looking for feedback on which fields should be required vs optional.***

OAE Metadata			
Fields		*Input descriptions for each metadata field are provided below, inputs shown in [brackets] should match the exact vocabulary given.*	*Illustrative example (not based on a real study).*
Project ID:		The project to which the submitted data belong. A unique project identifier that can be used to link project data across data submissions, and link baseline data to intervention data, for example. If no Project ID has been assigned, one may be generated by combining the following fields, as described in *Cross-linking Data Sets with Common Identifiers*. Any method that creates a unique ID that will link all project data (e.g., a project’s baseline data to intervention data, and various data	- Lead organizer: Carbon Dive - Project beginning August 05, 2025 - Hvalfjordur, Iceland *Carbondive_20250805_Hvalfjordur*

			submissions within an *experiment type*) is acceptable.	
134d	**Experiment type:**		[natural] [manipulated] [model output] *Note to reviewers: Other options for 'manipulated' might include 'adjusted', 'modified', 'engineered'. Please provide your feedback on these controlled vocabularies.* See *Controlled Vocabularies* section for definitions. For model generated data, move to *Model Description Template*.	manipulated
134e	**Experiment ID:**		The experiment to which the data belong. The recommended naming convention is: - Project ID - Project condition - Numerical indicator to differentiate between various experiments of the same type for a project. A two digit consecutive number beginning with 01	- First intervention (experiment type = manipulated) for this project *Carbondive_20250805_Hvalfjordur_manipulated01*
134f	**Observation type:**		Please select any of the following that describe the submitted data: [profile] [surface	field experiments

			underway] [time series] [laboratory experiments] [pelagic mesocosms] [benthic mesocosms] [field experiments] [natural analogues] See *Controlled Vocabularies* section for *Observation type* definitions. Model output metadata should use the *Model Description Template*.	
134g	**mCDR Pathways:**		[ocean alkalinity enhancement] [macroalgal cultivation for CDR] [direct ocean capture] [ocean fertilization] [artificial upwelling and downwelling] [coastal blue carbon] [marine ecosystem recovery] See *Controlled Vocabularies* section for definitions.	ocean alkalinity enhancement
134h	**Site description:**		Provide information to help characterize the field site and provide context when interpreting the data. For example, descriptions of tidal patterns, climatological conditions, notable geological characteristics, the geographical and marine setting (coastal, intertidal, island region, sheltered environment),	**Hvalfjordur, Iceland** The proposed field site is in Hvalfjordur, Iceland. The fjord is approximately 35 km long, 3.5 km wide and 15 - 50 m deep. The site has a sheltered physical environment with predictable circulation and water residence time. The flow in the fjord is

		local sentiment surrounding coastal activities and climate change, notable events that may impact local sentiment to mCDR (for example: site had significant toxic spill in past decade, local positive support for offshore wind farming, frequent HAB site), ecologically protected species, economically significant operations in the marine environment, characteristic meteorological events. If possible based on the file type of this submission, please include useful maps or figures here. Links to relevant social science surveys, cruise reports, or protected species observer notes may be provided here.	characterized by inflow at depth and outflow at the surface, with primarily counterclockwise circulation. Water temp ranges from 0° C in winter and 10° C in summer. Hvalfjordur experiences a subpolar oceanic climate characterized by strong downslope winds, increased rainfall due to its fjord-mountain landscape, and maritime temperature moderation from the North Atlantic current. The local community is represented in project governance (board) and is engaged actively via town halls, information sessions, a website and newsletter. Cultural activities in the fjord include mussel harvesting, though the toxicity of the mussels is monitored by the food agency in Iceland and is not always permitted. Global Attitude Surveys were conducted in 2010 and 2020 and reported in the Bjarnadóttir et al report *What do Icelanders think about the environment and climate change?* Economic activity in the fjord includes several

				areas zoned for sediment mining, a port in Grundertangi and an aluminum smelting plant, among others. Usually no pelagic fish in fishable quantities with the exception of winter 1947 - 1948 where large schools of summer spawning herring led to high catches. Moderate harbour seal population has rebounded in recent years.
134i	**Project description:**		A narrative description of the project. For example, what were the goals of the project? What were the research questions? What were the processes to achieve these goals and answer these questions? Who were the key stakeholders, organizers, project leaders? Was this building off a previous or ongoing project, or is this a new region/experiment/mechanistic study?	**Hvalfjordur MRV System Pilot Study** A baseline study beginning in 2024 captured the physical, biogeochemical, atmospheric and biological data of the site over the course of a year. It included autonomous and vessel-based samples as well as public data sources. Building on this year-long baseline study, Dual Tracer study, and Dye Tracer Study, an interdisciplinary project team under the leadership of Dr. Jane Doe conducted research to establish and test a prototype MRV System Pilot. The

				research questions were: 1) Can adding NaOH effectively increase seawater alkalinity, 2) Can increased alkalinity reduce surface ocean pCO_2, 3) Does reducing surface ocean pCO_2 result in CDR, and 4) is there an impact on local species and natural communities as a result?
134j	**Experiment description:**		A narrative description of the experiment. For example, what part of the project do these data represent (e.g, baseline, intervention, control) and what do they contribute to the overall project? Are all project research questions listed in *Project description* relevant? What were the processes to achieve these goals and answer these questions?	On August 10, 2025 the project team released 23 tons of diluted NaOH solution over 96 hours and observed the results for 14 days. This experiment represented the first intervention conducted in the region and for this project. All project research questions above are relevant as the intervention will allow these questions to be answered. To effectively monitor these study regions in order to answer these questions, 10 repeat ship surveys were conducted to collect grab samples, underway sensor data, and profile data from rosettes. 12 buoys were deployed with sensor arrays including

			temperature, salinity, oxygen, chlorophyll, particulate information, pH, and pCO_2. These data will be used to monitor the change in seawater pCO_2 and local species impact due to dissolved alkalinity addition.
134k **Other datasets collected from this project:**		If multiple data sets were produced out of the same experiment, please list all related unique identifiers (e.g., DOI's) to these datasets. Also include links to additional experimental data produced from this Project ID.	All data for the current experiment (*CarbonModel_20250805_Hvalfjordur_manipulated01*) are provided in this submission, with the current draft doi: http://doi.org/xx.xxx Additional experimental project data include: Baseline study: *Carbondive_20250805_Hvalfjordur_natural01* http://doi.org/xx.xxx Dye Tracer Study: *CarbonModel_20250805_Hvalfjordur_manipulated02* http://doi.org/xx.xxx Counterfactual model run: *CarbonModel_01012024_iceland_modeloutput01* http://doi.org/xx.xxx Earlier biological research relevant to mCDR operations can

				be found via the Marine and Freshwater Research Institute.
134l	**Previous mCDR research in the area:**		If previous or on-going mCDR field operations have occurred in the study domain by any project developer, they may be mentioned here either as a description, and/or if a reference to the study exists in the form of a data set, publication, etc, the DOI or other identifying information should be provided. Please provide direct links to data when available.	mCDR company Algae Lock was headquartered near the fjord and conducted some proprietary carbonate chemistry and algae farming research in the fjord.
134m	**Co-located operations:**		A description of any nearby operations that may influence the waters over the time period covered by this data. This might be a nearby mCDR project, a facility that discharges water with different characteristics than the inflow (e.g., a desalination plant), frequent boating operations, etc.	Aluminum smelting plant co-located in fjord, activities and plans unknown.
134n	**Data conflicts and unreported data:**		If data exist that are not provided by the project due to conflicts (e.g., geopolitical or other), data availability (e.g., a dataset is no longer available), it may be noted here.	Data from a citizen-based water quality effort were available between the years 2021 - 2023 and are informed by Carbon Dive project planning,

				but are no longer accessible.
134o	**Meteorological and tidal conditions:**		Include links to relevant datasets and/or a narrative description.	Wind data: Vindatlas.vedur.is Bathymetry data: Coast Guard Data from Atlas.lmi.is Land and water usage map: Vefsja.is Tide & weather data (sea level, wind, air pressure, temperature, salinity: Vedur.mogt.is
134p	**Permit number:**		Associated permit number(s).	permit #XYZ (permit pending, example is illustrative)
134q	**Permit approval document:**		Link to permit or document reference.	permit #XYZ (permit pending, example is illustrative)
134r	**Permitting authority:**		Name of organization or authority related to permitting, if applicable.	Ministry for Foreign Affairs (Utanríkisráðuneytið) Environmental Agency of Iceland (Umhverfisstofnun)
134s	**Additional details:**		Open text area to include additional information. These may include information for sediment processes data, biological data, or any other required information if not included in the main metadata or data files; see *Guidelines for Data Submission* for relevant sections of your data. Additional informational	See https://samplewebsite.is/data for a field blog and additional data from this site.

		files, such as digitized laboratory notebooks, blogs, etc., may be linked here.	

134t	Intervention Only:			
134u	**OAE treatment type:**		[electrochemical alkalinity addition] [mineral alkalinity addition] [dissolved alkalinity addition] [river alkalinity addition] [coastal enhanced weathering] See *Controlled Vocabularies* section for definitions.	dissolved alkalinity addition
134v	**Alkalinity feedstock type:**		[olivine] [potassium hydroxide] [lime] [portlandite] [calcium carbonate] [anorthite] [dolomite] [periclase] [brucite] [magnesite] [forsterite] [sodium hydroxide] [natrite] [nahcolite] See *Controlled Vocabularies* section for definitions.	sodium hydroxide
134w	**Alkalinity feedstock description:**		Information such as feedstock source, characteristics, concentration, impurities, dilution prior to dosing, and for feedstock other than NaOH; trace metal	30% NaOH solution (commercially acquired) mixed with freshwater to achieve 1050 kg/m^3 density. Tagged with 32g of inert gas SF_6, dissolved in 1000 liters of freshwater

		composition and particulate grain size.	
134x **Dosing location:**		For point source location, provide latitude and longitude in decimal degrees. For coastal dispersal, provide the maximum and minimum latitude and longitude positions. If the location varies, such as a ship deployment, provide the location as a latitude and longitude vector with the dosing data and time stamps, with column headers named 'intervention_location_lon' and 'intervention_location_lat' and state 'dosing location is provided as a variable' here. Latitude and longitude should be provided in decimal degrees (negative for southern and western hemispheres). In some cases, the dispersal may vary by only a few meters and a point location may be provided, however, in all cases, the decimal accuracy provided here should reflect the accuracy of known position.	64.394426, -21.465808
134y **Alkalinity dosing effluent density**:		Fixed density or provide link/source to effluent density data if applicable. Please	1050 kg/m^3 density, measured directly

			include whether density is directly measured or a derived value. If this is a variable included with your data, please note so here as ‘alkalinity dosing effluent density is provided as a variable’ and use ‘dosing_effluent_density’ for your column header name.	
134z	**Dosing depth(s):**		Depth(s) in meters. If this is variable, please include the schedule of depth changes and depths, or as a vector in meters with the data, named ‘dosing_depth’. Please note here that ‘dosing depth is provided as a variable’.	3 meters below surface
134aa	**Dosing mechanism:**		E.g., point source, outflow from pipe, coastal distribution, pier-based diffuser. Please be descriptive.	Pier-based diffuser
134ab	**Dosing regimen:**		At a minimum, please provide the schedule and timeline of dosing, including the time between doses, the duration of treatment and the amount used each time. More optimally, this information would be provided as a vector of binary data in the data file where 1 = dosing ‘on’ 0 = dosing ‘off’, using the	August 5, 2025: 3 IBC Test 09:00 - 12:00 August 6, 2025: 7 IBC dosing 09:00 - 16:00 August 7, 2025: 13 IBC dosing 09:00 - 22:00 August 8, 2025: 7 IBC dosing 09:00 - 16:00 Dosing regimen is also provided as a variable.

			column header name 'dosing_onoff'. If provided as a vector state here as 'dosing regimen is provided as a variable'.	
134ac	**Dosing rate data:**		To link dosing rate data that are provided a vector data (e.g., 'flow_rate', 'mineral_mass_addition', 'mineral_mass_addition_rate', or any of the fields above), provide the source or filename (if provided with the current data submission) to access these variables. If these are included in the data file, that should be stated here with the associated column header names listed.	All dosing rate data are provided in the current data submission in file: Carbondive_20250805_Hvalfjordur_manipulated01_dosingdata.csv Variables include: flow_rate (L/s)

135

135a	Physiological Response Studies:		
135b	**Targeted acidity or alkalinity levels:**	Description of the alkalinity range targeted.	NA
135c	**Manipulation method:**	A description of the manipulation method that was used. Associated DOIs can be noted here, but should be supplemental.	NA
135d	**Location where the experiment was carried out:**	City, state, and research facility name in which the experiment took place.	NA
135e	**Treatment duration:**	Start and end date and time (UTC)	NA

Model Description Template

The following template outlines key fields and examples for documenting model and model output data details in OAE research. It provides a structured framework to ensure clarity, consistency, and completeness in describing model components, grid details, boundary conditions, and project-specific protocols. This standardized approach supports the required model reproducibility goals of the Protocol.

Note to reviewers: Our goal is to ensure model runs can be as reproducible as possible. Please review for missing fields, conflicts with vocabulary or areas of unnecessary/overly granular detail. ***We are also looking for feedback on which fields should be required vs optional.***

Model Description Template		
Fields	*Input descriptions for each metadata field, inputs shown in [brackets] should match the exact vocabulary given.*	*Illustrative example (not a real project).*
Project ID:	The project to which the submitted data belong. A unique project identifier that can be used to link project data across data submissions and link baseline data to intervention data, for example. If no Project ID has been assigned, one may be generated by combining the following fields, as described in [*Cross-linking Data Sets with Common Identifiers*](). Any method that creates a unique ID that will link all project data (e.g., a project's baseline data to intervention data, and various data submissions within an *experiment type*) is acceptable.	- Lead organizer: CarbonModel - Project beginning January 01, 2024 - Iceland *CarbonModel_01012024_iceland*
Experiment type:	[natural] [manipulated] [model output]	Model output

		See *Controlled Vocabularies* section for definitions.	
147e	**Model run experiment type:**	[counterfactual] [perturbation] [counterfactual and perturbation] [other] See *Controlled Vocabularies* section for definitions.	perturbation
147f	**Experiment ID:**	The experiment to which the data belong. The recommended naming convention is: - Project ID - Experiment type - Numerical indicator to differentiate between various experiments of the same type for a project. A two digit consecutive number beginning with 01.	First model alkalinity run for this project *CarbonModel_01012024_iceland_modeloutput01*
147g	**Experiment description:**	A narrative description of the experiment. For example, what were the research questions and goals? What were the processes to achieve these goals and answer these questions?	The goals of this model run are to support an OAE field trial in Iceland by Carbon Dive in order to model the spread of alkalinity released and air-sea CO_2 exchange over time in the region.
147h	**Other datasets collected from this project:**	Include links or DOI's to associated field data and model runs for this Project ID.	Counterfactual model run: *CarbonModel_01012024_iceland_modeloutput01* http://doi.org/xx.xxx Baseline field study: *Carbondive_20250805_Hvalfjordur_natural01* http://doi.org/xx.xxx Dye Tracer Study: *CarbonModel_20250805_Hvalfjordur_manipulated02* http://doi.org/xx.xxx

147i	**Model configuration:**	Links to model configuration files (e.g. roms_application.h, roms.in, and build_roms.sh files for a ROMS simulation)	https://github.com/parkermac/LO_roms_user/tree/main/upwelling
147j	Model Physics Component		
147k	**Name:**	Name of model (e.g. ROMS, Oceananigans)	ucla-roms
147l	**Version:**	Model release version	tag-1
147m	**Codebase:**	Link to model code repository	https://github.com/CESR-lab/ucla-roms
147n	**Description:**	A description of the physical model characteristics, including version of equations being solved (hydrostatic vs non-hydrostatic), tracer advection scheme, how bottom drag is represented, mixed layer parameterizations, sub-grid mixing parameterizations if applicable, etc. Associated links to data, DOIs, or publications can be noted here, but should be supplemental.	The circulation model is a regional implementation of the Regional Ocean Modelling System (ucla-roms) Configured for the North Atlantic, centered on Iceland. The outer grid has a 3.3 km horizontal resolution and 100 vertical layers, while an inner nested grid has 40 m resolution and 100 vertical layers. ROMS is a free-surface, terrain-following, primitive equations ocean model, the hydrostatic primitive equations for momentum are solved using a split-explicit time-stepping scheme. All 2D and 3D equations are time-discretized using a third-order accurate predictor (Leap-Frog) and corrector (Adams-Molton) time-stepping algorithm. The primitive equations are discretized over variable topography using stretched terrain-following coordinates. The circulation model uses the 3rd-order upstream-biased (horizontal) and 4th-order centered differences (vertical) advection schemes for temperature and

			salinity. The model includes 12 freshwater inputs and is forced by the ERA5 atmospheric product (https://doi.org/10.1002/qj.3803) at the surface and by GLORYS at the boundaries. Vertical mixing is parameterized using the K-profile parameterization (KPP) from Large et al. 1994, and the air-sea interaction boundary layer in ROMS is based on the bulk parameterization of Fairall et al. (1996). Bathymetry is from GLORYS, the model T and S are initialized from GLORYS, and the model includes tides from the TPXO atlas.
147o	**References:**	Reference for model physics description	https://doi.org/10.1016/j.ocemod.2004.08.002
147p	Model BGC/Ecosystem Component		
147q	**Name:**	Name of BGC/Ecosystem component	MARBL
147r	**Version:**	Version of BGC/Ecosystem component used	Cesm2.2-n00 (imbedded in C-Star)
147s	**Codebase:**	Url link to where code can be found, the link to the specific commit (GitHub) or version should be provided.	https://github.com/marbl-ecosys/MARBL.git
147t	**Description:**	A description of the biogeochemical/biological model characteristics, including which parameters are modeled explicitly, derived carbonate system parameters, advection scheme for biological tracers, CO_2 solver protocol (e.g., CO2SYS), links to data/code with biological model parameters (e.g., growth and mortality rates), etc. Include details on whether	The Marine Biogeochemistry Library (MARBL) is a prognostic ocean biogeochemistry model that simulates marine ecosystem dynamics and the coupled cycles of carbon, nitrogen, phosphorus, iron, silicon, and oxygen and is a component of the Community System Earth Model 2 (CESM2). The ecosystem includes multiple phytoplankton

	dissolution and precipitation of calcium carbonate are considered, how exchanges between sediment and overlying water are represented (if applicable), and whether active feedbacks between biological processes and the carbonate system are represented. Associated links to data, DOIs, or publications can be noted here, but should be supplemental.	functional groups (diatoms, diazotrophs, small phytoplankton, and coccolithophores) and multiple potentially growth limiting nutrients (nitrate, ammonium, phosphate, silicate, and iron. There is one zooplankton group, dissolved organic material (semi-labile), sinking particulate pools and explicit simulation of the biogeochemical cycling of key elements (C, N, P, Fe, Si, O, plus alkalinity) (Moore et al. 2004). The ecosystem component is coupled with a carbonate chemistry module based on the Ocean Carbon Model Intercomparison Project (OCMIP) (Doney et al. 2009) allowing dynamic computation of surface ocean pCO2 and air-sea CO_2 flux. Photoadaptation is calculated as a variable phytoplankton ratio of chlorophyll to nitrogen based on Geider et al. 1998. Phytoplankton N/P ratios are fixed at the Redfield value of 16, but the diazotroph group has a higher N/P atomic ratio of 50. The model parameterizes a prognostic phytoplankton calcifier in MARBL that is modeled on coccolithophore physiology (Krumhardt et al., 2019). The ratio of calcification to photosynthesis by the coccolithophore functional type is responsive to environmental conditions, where the calcification to photosynthesis ratio is a function of temperature, nutrients, and

		CO_2. Carbonate chemistry is explicit and there are two parallel carbonate systems including DIC and alkalinity tracers; applying fixed-preindustrial and time-evolving atmospheric CO_2 to these parallel systems enables cleanly computing anthropogenic CO_2 concentrations. MARBL computes burial and denitrification losses of material at the seafloor according to empirical relationships. Particulate organic carbon burial is computed using a relationship between burial efficiency and POC flux from Dunne et al. (2007), with an imposed maximum burial efficiency of 80%. Burial of SiO_2 at the seafloor is based on observations in Ragueneau et al. (2000). In MARBL, 4% of Si incidents on the seafloor are buried, except where the incident flux of Si to the seafloor exceeds 2 mmol m−2 d−1; then, 20% of Si is buried. As described above, sedimentary denitrification depends on the incident POC flux and is computed based on an empirical relationship from Bohlen et al. (2012). Burial of $CaCO_3$ on the ocean floor occurs where $\Omega > \Omega crit$ in the model's bottom layer; where $\Omega < \Omega crit$, all $CaCO_3$ reaching the model's bottom layer is dissolved. All $CaCO_3$ is assumed to be calcite, thus ignoring the distinction between

			the mineral forms calcite and aragonite, which may be important in modulating dissolution depths (Gangstø et al., 2008).
147u	**Air-sea CO2 flux parameterization:**	Description and/or references of air-sea CO_2 flux parameterization used, gas transfer velocity formulation and atmospheric CO_2 details (e.g., fixed or time varying, and if time varying which data were used).	Air-sea CO2 gas exchange is parameterised as a function of temperature (T) and wind speed (u10), and the concentration of the gas in the air (Ca) and in the surface water (Cw) in the form: F = k(u10 T)(Cw-Ca), where k is the gas transfer velocity. Gas transfer velocity is parameterized using the 4th order polynomial formulation of Wanninkhof (2014). Quadratic k_{660} parameterisation is calibrated to give 16.5 cm/hr global average (recommended Naegler, 2009) for the ERA5 wind product by SeaFlux/Luke Gregor et al. (2023). Atmospheric CO_2 is assumed fixed and spatially uniform at 428 ppm.
147v	**References:**	Links or DOIs to any reference(s) relevant to the model components/development, specific model configuration, model validation etc.	https://doi.org/10.1029/2021MS002647
147w	**Other model components:**	Additional components such as sea ice, sediment, atmosphere, etc., following the structure above if applicable (e.g., Name, Version, Codebase, Description, References).	Not applicable
147x	Grid Details:		

147y	**Grid type:**	Descriptive structure of grid (e.g., latitude-longitude grid, unstructured triangular, tripolar)	Rectangular x-y grid with rotation: Central longitude: -19 Central latitude: 65 Grid Rotation: 20
147z	**Region:**	A description of the region modeled, include the latitude and longitude boundaries if they are not provided in another field. Latitude and longitude should be provided in decimal degrees (negative for southern and western hemispheres).	North Atlantic centered on Iceland Longitude range = -17° to -20° Latitude range = 60° to 70°
147aa	**Arrangement:**	The grid arrangement of orthogonal physical quantities (e.g. Arakawa A, Arakawa B, Arakawa C)	Arakawa C-grid
147ab	**Nx:**	Number of x grid points	800
147ac	**Ny:**	Number of y grid points	800
147ad	**Nz:**	Number of vertical coordinate levels	100
147ae	**N nodes:**	Number of grid nodes (if using an unstructured grid)	5285
147af	**Horizontal resolution range:**	Range of horizontal resolution (m or km)	3.3 km (for the outer nest)
147ag	**Vertical resolution range:**	Range of vertical resolution (m)	Max. 4 m (topography following vertical grid)
147ah	Input Details:		
147ai	**Bathymetry:**	Data source for bathymetry used (including links to data if available)	GLORYS bathymetry data- https://doi.org/10.48670/moi-00021
147aj	**Initial conditions**	Data sources for initial conditions of all model state variables (including links to data if available)	Initial conditions from GLORYS (mercatorglorys12v1_gl12_mean)
147ak	**Boundary conditions:**	Data source for boundary conditions for all model state	GLORYS (mercatorglorys12v1_gl12_mea

		variables (including links to data if available)	n)
147am	**Atmospheric forcing data used:**	Data source for atmospheric forcing if applicable (including links to data if available). Examples include, wind fields, shortwave and longwave radiation, air temperature, humidity, etc.	ERA5 hourly (https://doi.org/10.24381/cds.143582cf)
147an	**Tidal forcing:**	Data source for tidal forcing (including links to data if available)	TPXO atlas (https://www.tpxo.net/global/tpxo10-atlas)
147ao	**River & sediment flux details:**	Description of river and sediment flux data used to force the model (including links to data if available)	River fluxes for the inner nest sourced from the Icelandic Met Office (https://en.vedur.is/) for 12 rivers, no river fluxes used for the outer nest. No sediment fluxes applied.
147ap	Experiment Details:		
147aq	**Spin-up protocol:**	A description of the spin up process chosen for the model initiation, including an explanation for how appropriate spin up was defined to be achieved.	2 weeks per nest
147ar	**Start date and time:**	Start date and time of model experiment in UTC	01-Jan-2024 01:20:30
147as	**End date and time:**	End date and time of model experiment in UTC	01-Jan-2034 01:20:30
147at	**Output frequency:**	Time frequency at which model fields are saved (e.g. hourly mean, daily mean)	Monthly means
147au	**Time stepping scheme:**	Method used to discretize time domain (e.g., Euler, Runge-Kutta, leapfrog)	10 second for spin-up - up to 3 minutes per timestep for outer nest
147av	**Description of alkalinity addition:**	A description of how alkalinity perturbation was applied in the model	Applied over multiple grid cells in initial conditions to ALK_ALT_CO2 variable in MARBL (only in inner nest, no Alk experiment in outer nest)

Hardware Configuration:		
Machine:	Machine name of hardware used to run model	Perlmutter
Operating system:	Operating system of hardware used to run model	Linux
CPU/GPU details:	Details on CPU or GPU hardware	Details here: https://docs.nersc.gov/systems/perlmutter/architecture/#cpu-nodes
Memory:	Memory capacity of machine	512 GB of DDR4 memory total
Storage:	Storage capacity of operating system	44 PB
Parallelization:	Description of processors used in parallel, including number or processors and MPI version if used.	3 nodes and 108 ntasks per node

Guidelines for Data Submission

This section outlines the specific requirements and recommendations for submitting data associated with OAE research. It covers general guidelines for adjusted and raw data, in situ sensor data, sediment processes, and biological/physiological data. Additionally, it provides instructions for creating unique Project IDs to facilitate cross-linking of datasets, particularly for research cruises and other projects, and timelines for archiving data. These standards ensure data consistency, reproducibility, and compatibility across studies while allowing for flexibility to include supplementary details in the *OAE Metadata* when repository fields are insufficient.

In this section, a 🚩 indicates requirements in order to be considered compliant with this protocol. “Recommendations” are strongly suggested best practices to support quality data sharing, but may not always be possible.

General Guidelines for your Data

The following recommendations and requirements are general for both discrete and sensor data.

Requirements:

- 🚩 When preparing your data, refer to the data file templates provided in *Column Header Names* that include recommended column header names and units for various data types.

- 🚩 When filling out your metadata, refer to the *Controlled Vocabularies* for inputs.

- Field data representing baseline conditions, control conditions, or intervention conditions will be uploaded separately.

- The unique project ID must be included in any data upload's metadata in order to link e.g., the model output to other affiliated project data, baseline data to intervention data, etc.

- A unique sample identification number should be assigned to vector data (use column header name = 'sample_ID'). These can simply be consecutive numbers.

- If not included in the metadata (e.g., see fields under Observed Properties in NOAA's OCADS metadata file), a *ReadMe* file should be provided in plain text format. Contents may include the following as text, figures, charts, etc. (as applicable):
  - Description of calibrations or adjustments performed to data.
  - Any information needed to reproduce adjusted data or derived variables. This may include, but is not limited to, equations, coefficients (e.g., if using CO2SYS for carbonate parameter calculations; see *Derived Variables*).
  - References on data QA/QC processing steps (e.g., DOI, software and software settings used, etc).
  - For sensor data, include the sensor ID and/or serial number, certification number, as available.
  - Analysis protocols should be described, or a DOI provided.

- Missing values should be reported as '-999'.

- Associated quality flags for measured variables should be indicated as <parameter_name>_flag and follow the following flagging definitions (Jiang et al., 2022; Table 2),
  0 = interpolated or calculated data
  1 = not evaluated/quality unknown
  2 = acceptable
  3 = questionable
  4 = known bad
  6 = median of replicates
  9 = missing value

Recommendations:

+ If processing software were used, include firmware version in *ReadMe*.
+ Sampling protocols should be listed in the *ReadMe* file.
+ Providing meteorological data that provides context to the results of the project and interpretation of data is strongly encouraged. For example, rain accumulation (as it impacts salinity).

+ Point by point uncertainty/precision values should be included with data (indicated as <parameter_name>_uncertainty or <parameter_name>_precision). For sensor data, if no uncertainty or precision analyses were performed during the project, these may be estimated using the manufacturers calibration sheets, for example, a factory calibrated sensor may have an accuracy specification of "+/-2 uM or 1.5% of measurement, whichever is greater." This information can be converted into a point-by-point uncertainty value.

Guidelines for In Situ Sensor Data

Requirements:

- If adjusted data are provided, raw data must be provided for sensor and discrete data (indicated as <parameter_name>_raw and <parameter_name>_adjusted).

- Manufacturer calibration files are required with submission, if the sensor ID is not included, it should be included in the metadata or *ReadMe* file.

Recommendations:

+ It is strongly recommended that sensor data are provided in netCDF, especially when reporting data from a multi-sensor platform.

+ Profile data may have associated standard deviation profile or median profile values (indicated as <parameter_name>_std and <parameter_name>_med).

Guidelines for Model Output

Model output data plays a crucial role in interpreting and predicting the outcomes of OAE interventions. Adhering to established data standards ensures that model configurations and outputs are consistent, interoperable, and easily integrated with observational data. This section outlines the requirements to maximize the usability and reproducibility of model output, as well as to facilitate meaningful comparisons across studies and datasets.

Requirements:

- All model code, including software versions (e.g., compiler information), model configuration, experiment codes and input/forcing data wherever possible must be shared publicly. Given data storage costs and limitations, it is not expected that large input datasets (e.g. atmospheric forcing data or boundary conditions) that are publicly available elsewhere be archived along with model code, but model code should link to public data repositories containing forcing. If model forcing datasets are not already stored publicly, it is recommended that they be publicly archived and linked to the model code wherever possible.

- Documentation outlining the model components, code versions, configuration, forcing, and parameterizations named 'model_description' if uploaded separately and provided in, xlsx, netCDF, or plain text format. An example template for this information is provided in the *Model Description Template*.

- Parameter names may refer to *Column Header Names and Controlled Vocabularies* for common naming conventions.

- The parameters modeled, spatial, and temporal frequency of model output data will vary substantially for specific model projects, and archiving large quantities of model output is impractical due to data storage limitations and costs. However, in cases where a biogeochemical model is used to quantify net CDR (using a baseline and intervention experiment), we recommend publicly archiving the following parameters from models from both baseline and intervention model experiments at minimum:

  2D Air-sea CO_2 flux (mol/m^2/s or mmol/m^2/day) time series
  3D Dissolved Inorganic Carbon (µmol/kg) time series
  3D Total Alkalinity (µmol/kg) time series
  3D Temperature (°C) time series
  3D Salinity time series

  To save space, derived variables do not need to be included, however the inputs needed to derive those variables (e.g. for CO2SYS) should be included. It is recommended that model output is archived on the original model grid, but the temporal frequency of the provided data is project dependent and left to the data provider to decide what to supply.

Recommendations:

+ The data from a counterfactual run and the intervention run may each be provided separately, as this can decrease the data storage needed.

+ It is recommended that model outputs are provided in netCDF format, following recommended naming conventions (see *Column Header Names and Controlled Vocabularies* for parameter naming conventions).

+ Additional recommended parameters to archive include:
  - Physical transport vectors (u, v, w)
  - Biological tracers (chlorophyll, phytoplankton, zooplankton, dissolved oxygen, NO_3^- and NH_4^+, as available)

Guidelines for Intervention Data

An intervention refers to the deliberate action taken to alter ocean chemistry, such as adding alkalinity or reducing acidity, as part of an Ocean Alkalinity Enhancement (OAE) study. This section builds on the metadata requirements by detailing the additional intervention-specific data needed to fully document and contextualize the action, ensuring consistency and utility across studies. The following Requirements and Recommendations are given for intervention data, in addition to those listed under *General Guidelines for your Data*.

Requirements:

- Data related to dosing must be included*, for example:
    - Flow rate
    - Mineral mass addition
    - Mineral mass addition rate

  *These can be detailed in a separate file, with associated links provided in the *OAE Metadata*, or as data vectors (see Recommendations below).

Recommendations:

+ Providing the dosing information above as a vector, interpolated to the intervention data is suggested, using variable names: 'flow_rate', 'mineral_mass_addition', and 'mineral_mass_addition_rate'.

+ For usability, a binary vector describing whether the dosing is on/off is recommended with the data submission, e.g.,
  'dosing_onoff' : *0 = no dosing is occurring, 1 = dosing is active.*

Guidelines for Sediment Processes Data

Requirements:

- A description and/or reference to methods used for total carbon values must be included in the metadata. If there is no appropriate field to include this information in the data submitter's repository, it can be included in the OAE Supplemental Metadata file under the field 'Additional information'.

- Sediment sampling method should also be included. If there is no appropriate field to include this information in the data submitter's repository, it can be included in the OAE Metadata under the field 'Additional details'.

Recommendations:

+ Local sediment type should be included in the OAE Supplemental Metadata under 'Site description'.

+ Information on water depth at the sediment collection site and core length should be provided in the data file.

Guidelines for Biological/Physiological Data

Recommendations:

+ Information on time scales and species level should be provided. If there is no appropriate field to include this information in the data submitter's repository, it can be included in the OAE Metadata under the field 'Additional details'.

Derived Variables

Derived variables refer to those that are estimated from measured variables. These may be estimated utilizing software or additional calculations.

Carbonate Parameters

When using CO2Sys to derive carbonate parameters, the user has a choice of equilibrium constants. Please refer to the documentation in PyCO2Sys for recommended constants appropriate to your oceanographic conditions. A general recommendation for seawater with temperature between 2° and 35°, and salinity between 19 to 43, is to use the carbonic acid dissociation constants of Leuker et al. (2000) and Waters et al. (2014) for brackish water (with the salinity range 1 to 50 and temperature range 0 - 50). The boron-salinity ratio from Lee et al (2010) is recommended, and for the dissociation constants of bisulfate (HSO_4^-), Dickson (1990) is recommended. For the dissociation constants of hydrofluoric acid (HF), Perez and Fraga (1987) is recommended. See Jiang et al. (2022) for further discussion of these recommendations and associated uncertainties.

See fields under xCO_2/pCO_2/fCO_2 in NOAA's OCADS metadata file for additional required metadata fields for continuous data of calculated pCO_2/fCO_2.

TEOS-10 Calculations

Software to calculate common ocean properties (e.g., density, absolute salinity, potential temperature) based on thermodynamic equations can be downloaded for use with most coding languages here: https://www.teos-10.org/software.htm and for use with excel here: https://github.com/dpierrot/GSW_Sys.

Cross-linking Data Sets with Common Identifiers

Project ID

An ID that is unique based on the project's key attributes may be created using the following fields, separated by an underscore character, however any method that creates a unique ID that will link all project data (e.g., a project's baseline data to intervention data, and various data submissions within an experiment type) is acceptable.

- Lead organizer surname and first initial, or company
- A unique date with the primary purpose to ensure a unique Project ID, as such the guidelines here are not strict. Please choose a date most reasonable to the data, for example, this may be the date that data were first collected for this project (likely the

beginning of baseline sampling). Provide the date in Coordinated Universal Time (or UTC) using ISO-8601 format (YYYYMMDD)
- Location (descriptive)

Example 1: Carbon Dive is submitting project data beginning January 20, 2023 for their study site in San Francisco Bay where they collected baseline and intervention data, their resulting Project ID is **carbondive_20230120_sanfranciscobay**.

Experiment ID

An experiment ID is provided for each unique *Experiment type* as well as unique experiments within a single *Experiment type.* Options include using EXPOCODE (for research cruise projects), or generating one by combining the following:
- Project ID (see above for how to generate)
- Experiment type (see *Controlled Vocabularies*)
- A numerical indicator starting at 01 to differentiate between experiments of the same type for a project.

In the example for Carbon Dive above, their Project ID is carbondive_20240120_sanfranciscobay.

Carbon Dive collected baseline and intervention data during this project. For the baseline experiment, the Experiment ID is carbondive_20240120_sanfranciscobay_**natural01**, and the Experiment ID for their intervention is carbondive_20240120_sanfransiscobay_**manipulated01**.

They also conducted an intervention experiment that was deemed reasonably distinct from the first (e.g., the details under *OAE Metadata Supplement* for carbondive_sanfransiscobay_manipulated01 were not representative of this intervention experiment, however it was still considered to be under the same project), and used Experiment ID carbondive_20240120_sanfransiscobay_**manipulated02**.

Example 2: Dr. Jane Smith is providing model output from two global studies, one from a counterfactual model run and one from a baseline model run, their resulting Project ID is smith_20241004_global, and Experiment IDs are smith_20241004_global_**modeloutput01** and smith_20241004_global_**modeloutput02**

Guidelines for Archiving

Naming convention of data files:

It is recommended that individual data files are named by combining the Experiment ID with the data type. For example, a file containing data from discrete carbonate bottle samples during the first OAE intervention of project carbondive_20250805_Hvalfjordur will be titled carbondive_20250805_Hvalfjordur_manipulated01_**discrete**. If multiple discrete bottle files will

be uploaded, appending a numerical value to the end starting with 01 can differentiate these. However, it is recommended that data of a single type are uploaded together.
Other examples of data types include: profile (e.g., data files containing discrete and sensor data from a CTD cast), field (for field data files containing a mix of data types), autonomous, underway, discrete, physiological, and model output.

Timeline for data submission:
Deadlines for submitting data depend on the intended use of the data. For example, data submitted for regulatory purposes often required a faster turnaround to meet the demands of project development, while data for academic investigations may not face the same external pressures. However, the results of academic projects are fundamental to advancing mCDR safely and responsibly and have therefore been allocated appropriate timelines to reflect this importance.

Reporting and sharing model output may largely be dictated by project and registry timelines, and verifiers will need to verify model results prior to credit delivery, requiring models to be rigorously tested and validated before the intervention project begins to ensure that accurate model-based quantification of CDR can be executed and shared on a scale of weeks, depending on planned reporting periods.

Data for all purposes should be submitted by the earlier of the deadlines set by the project funders, regulators, registry, and other stakeholders, or the following:

Sensor data: Delayed-mode (e.g., quality controlled and adjusted) data should be made available within 3 months from collection (when the sensor is placed in-situ). It is recommended that real time sensor data be archived within 1 month.

Model output: Within 3 months of completion of model experiments.

Discrete data: Discrete data from samples including carbonates, sediment processes, nutrients, biological data etc. should be made available within 3 months from collection.

Where to store data:
Data can be stored in any scientific data repository that provides long-term preservation of data (ideally with version control capabilities), rich metadata management supporting the mCDR metadata template as outlined previously, and that provides data citations with a unique DOI. The choice of data repository may often be dictated by funder requirements. However, we make the following suggested recommendations for data repositories.

- Discrete and sensor data are recommended to be stored at NOAA's Ocean Carbon and Acidification Data System (OCADS).
- Raw eDNA data are suggested to be archived at the National Center for Biotechnology Information (NCBI).

Column Header Names and Controlled Vocabularies

A fundamental principle of data standardization is ensuring consistent naming across data submissions, projects, and repositories. If your chosen repository does not enforce a vocabulary, please refer to the references below for recommended naming conventions. Recommended units are also frequently provided for each variable, which significantly eases the process of synthesizing data across projects when standardization is maintained.

Column Header Names

Column header names are the labels given to specific data types, variables, or measurements in a data set, commonly presented as abbreviated names in column headers. Standardizing column header names significantly eases syntheses projects and ensures that a user can find data of interest. Data file templates are provided below that include recommended column header names. It is recommended to use the following data submission templates for the following data types (provided in .xlsx files), this will ensure column header names and units are standardized across projects.

Note to reviewers: Please select any of the following data types that you are most familiar with to provide feedback on the column header names for these data.

Discrete data:

- Column Header Names Description
- Data File (.XLSX) Example

Autonomous data:

- Column Header Names Description
- Data File (.XLSX) Example
- Further naming conventions of less-common and diagnostic variables are provided by Argo

Underway data:

- Column Header Names Description
- Data File (.XLSX) Example

Physiological response data:

- Column Header Names Description
- Data File (.XLSX) Example

Controlled Vocabularies

Controlled vocabularies are separate from column header names. These are controlled definitions of whole descriptive words that may be used for metadata input, for example.

Note to reviewers: the motivation of the controlled vocabularies for experiment type is to be able to pull out un-perturbed data from a large database (or in this example ‘natural’ data), for example. This may relate to mCDR experiments, but there are many experiments that would be considered ‘manipulated’ data. As such, this needs to be broad and not only mCDR-applicable (as is the case with many of the controlled lists that follow).

Controlled Vocabularies for Experiment Type	
Experiment type	*Definition*
Natural	Refers to measurements representing its natural environment or context. This includes gridded or calculated values. For example, the gridded SOCAT data product is considered an in-situ data product.
Manipulated	This refers to data or measurements representing conditions that have been adjusted, or transformed from their original state. For example, observations from mCDR field trial, mesocosm, laboratory experiment, are all considered manipulated values.
Model output	Model output refers to the results or data generated by numerical or computational models.

Controlled Vocabularies for mCDR Pathways	
mCDR Pathway	*Definition*
Ocean alkalinity enhancement	Ocean Alkalinity Enhancement (OAE) is a method to help mitigate climate change by increasing the alkalinity of seawater to enhance its capacity to absorb and store atmospheric carbon dioxide (CO_2).
Macroalgal cultivation for CDR	Macroalgal Cultivation refers to the process of farming large seaweeds (macroalgae) such as kelp, sargassum, or other marine plants to absorb atmospheric carbon dioxide (CO_2) and potentially sequester it over the long term.
Direct ocean capture	Direct Ocean Capture (DOC) is a technology-driven approach to extract carbon dioxide (CO_2) directly from seawater.
Ocean fertilization	Ocean Fertilization is a mCDR strategy that involves adding nutrients, such as iron, nitrogen, or phosphorus, to the ocean to stimulate the growth of phytoplankton or other microscopic plants that absorb carbon dioxide (CO_2) through photosynthesis.

Artificial upwelling and downwelling	Artificial Upwelling and Downwelling are mCDR strategies that involve manipulating ocean water movement to enhance natural carbon sequestration processes.
Coastal blue carbon	Coastal Blue Carbon refers to the carbon captured and stored by coastal ecosystems, such as mangroves, salt marshes, and seagrasses. These ecosystems absorb carbon dioxide (CO_2) from the atmosphere and store it in their biomass (leaves, roots, stems) and sediments, making them natural and effective solutions for mCDR.
Marine ecosystem recovery	Marine Ecosystem Recovery refers to the restoration and protection of marine ecosystems to enhance their natural ability to capture and store carbon dioxide (CO_2). This approach leverages the natural carbon-sequestering processes of marine habitats like coral reefs, kelp forests, seagrass meadows, oyster beds, and deep-sea ecosystems, aiming to rebuild biodiversity, ecosystem functions, and carbon storage capacity.

Note to reviewers: please review the suggested vocabulary list and definitions below. We aim to create easily recognizable categories of OAE treatment types while grouping activities that may be relevant for field data synthesis.

Controlled Vocabularies for OAE Treatment Types	
Treatment type	*Definition*
Electrochemical alkalinity addition	Uses electrochemistry to increase seawater alkalinity. This process typically involves separating seawater into acidic and alkaline streams, with the alkaline stream being released into the ocean to enhance carbon sequestration, with the acidic stream neutralized on land or used as a byproduct.
Mineral alkalinity addition	Involves adding alkaline minerals or particulate slurry (such as $MgOH_2$, MgO, or CaO) to seawater either directly or through coastal outfalls (such as wastewater) to increase its alkalinity.
Dissolved alkalinity addition	Involves the addition of dissolved or aqueous alkaline feedstocks (e.g. NaOH solution) into seawater either

	directly in coastal waters, through coastal outfalls or in the open ocean.
River alkalinity enhancement	Involves adding alkaline substances (such as limestone) in rivers before they flow into the ocean for the purposes of carbon removal or emission reduction.
Coastal enhanced weathering	Introduces alkaline mineral sand (e.g. olivine), boulders or berms in coastal areas to promote the absorption of CO_2 from the atmosphere into carbonate minerals (e.g., calcium carbonate, $CaCO_3$).
Pre-equilibrated alkalinity addition	Involves pre-treating seawater with alkalinity in a controlled setting, allowing equilibration with the atmosphere before returning into the marine system.

Controlled Vocabularies for Model Experiment Types	
Model Experiment Types	*Definition*
Counterfactual	A counterfactual model experiment describes a simulation created to mimic current oceanic conditions without interventions, such as mCDR treatment.
Perturbation	A perturbed model experiment describes a simulation created to mimic an intervention or change from the natural ocean conditions.

Mineral feedstock source materials for OAE (based on Renforth and Henderson, 2017; Caserini et al., 2022). This list includes input options to the ‘Alkalinity feedstock type’ field of the *OAE Metadata*.

Note to reviewers: The controlled vocabularies below are previously vetted and provided for reference only. (No additional review is needed)

Additional relevant vocabularies may include the following categories:
Observation types (Jiang et al., 2023b)
Instrument type (Jiang et al., 2023b)
Platform type (Jiang et al., 2023b)
Country: provided by the NERC vocabulary server
Sea names: provided by SeaDataNet C16 (sea areas)
Institutions: provided by the Research Organization Registry (ROR)

Research Vessel Platform Codes: provided by the International Council for the Exploration of the Sea (ICES)

Emerging Standards

This section describes in-work standards, file structures, etc., that are supplementary or complimentary to the OAE Data Management Protocol.

There are multiple areas that are currently under development or have not been covered within the current version of the protocol. However, work is ongoing to broaden the current protocol to be a more robust resource for the OAE community.

If you are interested in providing input or leadership on the development of data management guidelines for these emerging areas, please reach out to Irene Polnyi, Director of Strategy and Partnerships at Carbon to Sea Initiative (irene@carbontosea.org).

Content expanded in updated protocol versions may include:
Sensor quality control recommendations: Consensus on the best practices for sensor quality control and calibration will be documented, with the goal to provide a list of useful resources (guides, publications, tools, software, etc.), starting with the common biogeochemical sensors (nitrate, pH, pCO_2, dissolved oxygen, particle backscatter, chlorophyll fluorescence, and irradiance).

Biological experimental best practices: During the drafting of the current protocol, the OAE Data Standards Biological Working Group highlighted the need for standardized best practices in methods, calibrations, and sampling specific to OAE projects. Insights and outcomes from focused efforts in this area will be integrated into future iterations of the protocol.

Sediment processes data best practices: Work is in progress by the sediment processes community to standardize methodological approaches. Guidelines will be added to this protocol as their work develops.

Social sciences data best practices: Work is in progress by the social science community to develop further recommendations. Guidelines will be added to this protocol as these recommendations develop.

Laboratory and mesocosm studies: Although the current protocol focuses on field data, laboratory and mesocosm experiments play a crucial role in advancing OAE research. These data come with unique standardization challenges that need to be addressed. As recommendations become available, future iterations of this protocol will incorporate tailored guidelines for laboratory and mesocosm data, informed by input from experts in these areas.

Version control: A section on "recommendations for version control" that includes recommendations for updating data due to errors found in previous versions, updated methods, and/or bad data found will be discussed in future versions of the protocol.

Deprecated Standards and Practices

Previous versions of this protocol can be found on the host site.
The versioning format is defined as follows,
Update definitions:

MAJOR: Significant changes, such as new sections, substantial revisions, or breaking changes to existing recommendations.
MINOR: Additions or refinements to existing recommendations that are backward-compatible.
REVISION: Small edits like fixing typos, formatting, or minor clarifications without changing the meaning or scope.

Triggers for updates:
1. MAJOR Version (e.g., `2.0.0`):
- Introduced a new framework or restructured the document.
- Added entirely new priority levels or sections.
- Removed or deprecated previous recommendations in a way that could affect adoption.
- Example: Moving from draft to the first official release (`0.1` → `1.0`).

2. MINOR Version (e.g., `1.2.0`):
- Added new recommendations that complement existing ones.
- Updated examples, case studies, or appendices without altering the core content.
- Clarified ambiguities in key recommendations while maintaining backward compatibility.

3. REVISION (e.g., `1.1.3`):
- Corrected grammar, typos, or formatting.
- Reworded text for readability without changing intent.
- Fixed links or citations.

Acknowledgements

Coordination Team
The coordination team was responsible for the conceptualization and delivery of the protocol.

Leads: Jacki Long, LiQing Jiang, Veronica Tamsitt
With funding and steering support from: Irene Polnyi, Anna Madlener, David P. Keller

Data Initiative Steering Committee

The steering committee comprises members who are active in OAE field work to provide strategic guidance to the Data Management initiative coordination team as they assess the landscape of field data, surface areas of existing alignment and identify areas of opportunity.

Steering Committee: Gabby Kitch, David “Roo” Nicholson, Ruth Musgrave

Advisory and Consultation

Strategic guidance and consultation was provided from a range of experts via interview, document feedback and thought partnership.

Advisory and consultation: Jing He and Sophie Gill, Isometric; Tannis Thorlakson and Vilas Rao, Cascade Climate; Aaron Olson, Planetary Technologies; Jonathan Sharp, UW/PMEL.

OAE Data Dive Workshop Participants

Participants of the OAE Data Dive Workshop (Halifax, September 2024) provided valuable input into the first draft of the protocol, including thoughtful discussion around important trade-offs, surfacing key areas of alignment and driving towards internal consensus.

Grace Andrews, Hourglass Dariia Atamanchuk, Dalhousie Alice Benoit-Cattin, HAFRO Will Burt, Planetary Brendan Carter, NOAA/PMEL Manish Devana, Aquatic Brad Covey, Dalhousie Tim Dyson, Blue Sonde Jessica Foley, iSENSYS Carole Graas, IEEE Alyssa Griffin, UC Davis Chandler Griffin, iSENSYS David Ho, [C]Worthy Robert Izett, Planetary Alicia Karspeck, [C]Worthy Julie LaRoche, Dalhousie Matt Long, [C]Worthy Maike Luiken, IEEE	Alexandra Mackay, NOAA Andreas Macrander, HAFRO Enrico Milanese, WHOI Kate Morkeski, WHOI Daan Reijnders, Ebb Carbon Mallory Ringham, Ebb Carbon Carlos Muñoz Royo, AtDepth Laura Stieghorst, Básico Carbon Andrew Thompson, Blue Sonde Anya Waite, OFI Hongjie Wang, URI Liza Wright-Fairbanks, NOAA Casey Wilson, Subtidal

The Biological Working Group for OAE Data Standards

Keisha Bahr, Harte
Nina Bednarsek, USU
Brad Covey, Dalhousie
Jason Graff, Oregon State
Alyse Larkin, NOAA
Julie LaRoche, Dalhousie

Melissa Melendez, Uni. Hawaii
Shannon Meseck, NOAA
Charly Moras, Uni. Hamburg
Kai Schulz, SCU

The Sediment Working Group for OAE Data Standards

Brent Law, DFO-MPO
Grace Massey, VIMS
Ole Mikkelsen, Sequoia
Sara Nawaz, American Uni.
Christina Shultz, Northeastern

The Social Science Working Group for OAE Data Standards

Liliana Bastian, Ocean Visions
Meg Chadsey, Sea Grant
Brad Covey, Dalhousie
Terre Satterfield, UBC-IRES
Michael Weir, WHOI

AGU’24 OAE Data Management Guidelines Workshop participants

Anna-Adriana Anschütz, Leibniz Institute for Baltic Sea Research Warnemuende
Dariia Atamanchuk, Dalhousie University
Rick Berg, Banyu Carbon, Inc.
Kira Biener, Yale University
Eugene Burger, NOAA/OAR/PMEL
Will Burt, Planetary Tech
Mattias Cape, Environmental Defense Fund
Meg Chadsey, Washington Sea Grant & NOAA PMEL
Jessica Cross, DOE-PNNL
Suhani Dalal, Yale University
Nico Fairbairn, U.S. Senate (Schatz)
Luke Gloege, Yale University
Ryan Green, UC Santa Cruz / Oceanid Climate and Carbon Solutions
Alyssa Griffin, University of California, Davis/Bodega Marine Laboratory
Jing He, Isometric
David Ho, [C]Worthy
Diane Hoskins, Carbon to Sea
Linus Kamb, NOAA/Pacific Marine Environmental Laboratory
Gabby Kitch, NOAA OAP
Nicholas Kleinert, Carbon to Sea
Kristin Kleisner, EDF
Alyson Myers, Fearlessfund.org
Josh Perfetto, Ephemeral Carbon
Mallory Ringham, Ebb Carbon
Bita Sabbaghzadeh, Ephemeral Carbon
Jon Sharp, University of Washington CICOES / NOAA PMEL
Kirby Simon, Sequoia Scientific, Inc.
Laura Stieghorst, Básico Carbon
Nicolas Theunissen, Yale University
Andrew Thompson, bluesonde technologies Inc.
Jesse Vance, Ebb Carbon
Sally Walker, Applied Research Laboratories at UT Austin
Jennifer Yin, Isometric

Contributing Authors during the Open Comment Review Period

References

Dickson, A. G. (1990). Standard potential of the reaction: AgCl(s) + 0.5 H_2(g) = Ag(s) + HCl(aq), and the standard acidity constant of the ion HSO_4^- in synthetic sea water from 273.15 to 318.15 K. Journal of Chemical Thermodynamics 22, 113–127. doi:10.1016/0021-9614(90)90074-Z.

Frankignoulle, M., Canon, C., and Gattuso, J.-P. (1994). Marine calcification as a source of carbon dioxide: Positive feedback of increasing atmospheric CO_2. Limnology and Oceanography 39, 458–462. doi:10.4319/lo.1994.39.2.0458.

Jiang, L. Q., Kozyr, A., Relph, J.M. et al. The Ocean Carbon and Acidification Data System. Sci Data 10, 136 (2023). https://doi.org/10.1038/s41597-023-02042-0

Jiang, L. Q., Subhas, A., Basso, D., Fennel, K., and Gattuso, J.-P. (2023). Data reporting and sharing for ocean alkalinity enhancement research, in: Guide to Best Practices in Ocean Alkalinity Enhancement Research (OAE Guide 23), edited by: Oschlies, A., Stevenson, A., Bach, L., Fennel, K., Rickaby, R., Satterfield, T., Webb, R., and Gattuso, J.-P., Copernicus Publications, State Planet, https://doi.org/10.5194/sp-2-oae2023-13-2023.

Jiang, L. Q., Pierrot, D., Wanninkhof, R., et al., (2022). Best Practice Data Standards for Discrete Chemical Oceanographic Observations. Front. Mar. Sci, 8.

Jiang, L. Q., O'Connor, S. A., Arzayus, K. M., & Parsons, A. R. (2015). A metadata template for ocean acidification data, Earth Syst. Sci. Data, 7, 117–125.

Lee, K., Kim, T.-W., Byrne, R. H., Millero, F. J., Feely, R. A., and Liu, Y.-M. (2010). The universal ratio of boron to chlorinity for the North Pacific and North Atlantic oceans. Geochimica et Cosmochimica Acta 74, 1801–1811. doi:10.1016/j.gca.2009.12.027.

Lueker, T. J., Dickson, A. G., and Keeling, C. D. (2000). Ocean pCO_2 calculated from dissolved inorganic carbon, alkalinity, and equations for K1 and K2: validation based on laboratory measurements of CO_2 in gas and seawater at equilibrium. Marine Chemistry 70, 105–119. doi:10.1016/S0304-4203(00)00022-0.

Waters, J., Millero, F. J., and Woosley, R. J. (2014). Corrigendum to “The free proton concentration scale for seawater pH”, [MARCHE: 149 (2013) 8–22]. Marine Chemistry 165, 66–67. doi:10.1016/j.marchem.2014.07.004.

Appendix

Appendix A: Column header names for discrete data:

OCADS Column Header standards for Profile Data. Recommended column header abbreviations, recommended units, and brief descriptions, for discrete chemical oceanographic observations. In this table, CTD refers to the group of instruments for measuring conductivity (salinity), temperature, and depth, and CTD-rosette to the complete system of Niskin bottles (used for seawater sampling) on a frame together with the CTD. "N/A" means not applicable. "DP" is short for decimal places, or the number of digits after the decimal point. For more information, please check out Jiang et al., 2022.

Column Header standards for Profile Data			
Abbreviation [unit]	*Full unit*	*DP*	*Description*
EXPOCODE	N/A	N/A	Expedition code consists of the four-digit International Council for the Exploration of the Sea (ICES) platform code and the date of departure from port (UTC) in ISO8601 format (YYYYMMDD). For example, a research expedition onboard NOAA Ship Ronald H. Brown (ICES code: 33RO) leaving the port on August 27, 2015 (Coordinated Universal Time, or UTC) would have an EXPOCODE of 33RO20150827.
Cruise_ID	N/A	N/A	Cruise identification is the particular ship cruise number or other alias for the cruise. A Cruise_ID (e.g., A16N2013) could consist of a Section_ID (A16N) and the sampling year (e.g., 2013), if only one section is covered during the cruise.
Section_ID	N/A	N/A	Identification for a cruise section/leg (e.g., A16N, or P02).
Station_ID	N/A	N/A	Station identification. Numerical Station_IDs without letters are recommended to facilitate future QC efforts.

Cast_number	N/A	N/A	Cast number, where a cast is the lowering of equipment over the side at one station, e.g., CTD, net tow, etc. Cast_number should be sequential and restart with 1 for each station.
Rosette_position	N/A	N/A	Rosette position refers to the position number around the CTD-rosette (e.g., 1 of a 1-12, or 1-24, or 1-36 number).
Niskin_ID	N/A	N/A	Niskin_ID is a unique alphanumeric identifier assigned to only that Niskin bottle over the duration of the expedition.
Niskin_flag	N/A	N/A	Quality control flag for tracking problems with Niskin closure and integrity.
Sample_ID	N/A	N/A	A sample identifier (Sample_ID), which uniquely identifies a row of data during the subsequent QC and interpretation process, is often generated by concatenating the Station_ID, Cast_number, and Rosette_position, according to: Sample_ID = Station_ID × 10000 + Cast_number × 100 + Rosette_position. For example, at station 15, the 2nd cast, a Rosette_position of 3 will have a Sample_ID of 150203.
Year_UTC	N/A	0	Calendar year in UTC when Niskin bottles at a specific depth are triggered
Month_UTC	N/A	0	Calendar month in UTC when Niskin bottles at a specific depth are triggered
Day_UTC	N/A	0	Calendar day in UTC when Niskin bottles at a specific depth are triggered
Time_UTC	[hh:mm:ss]	N/A	Time in UTC (hh:mm:ss) when Niskin bottles at a specific depth are triggered
Yearday_UTC	N/A	2	Yearday refers to the day number in an annual cycle. (e.g., 06:00 on Jan 1 means yearday = 1.25, 18:00 on Dec 31 means yearday = 366.75 in a leap year). Note, Yearday_UTC starts with 1, instead of 0. It can be calculated according to this equation: Yearday_UTC = datefunction(Year_UTC, Month_UTC, Day_UTC)

			- datefunction(Year_UTC, 1, 1) + Time_UTC + 1, where, “datefunction” is the date function of a program (e.g., in Excel, the data function would be “DATE”).
Latitude	decimal degrees	4	Latitude in decimal degrees North (negative for southern hemisphere) when Niskin bottles at a specific depth are triggered
Longitude	decimal degrees	4	Longitude in decimal degrees East (negative for western hemisphere) when Niskin bottles at a specific depth are triggered
Depth_bottom	m	0	Bottom water depth of the sampling station
CTDPRES	dbar	1	Hydrostatic pressure recorded from CTD at the depth where the sample is taken
Depth	m	1	Depth at which a sample is taken. It can be approximated from CTDPRES and Latitude using the TEOS-10 equation.
CTDTEMP_ITS90	deg_C	3	*In situ* temperature recorded from CTD on the ITS-90 scale. If the temperature scale is IPTS-68, this term should be replaced with “CTDTEMP_IPTS68”.
CTDTEMP_flag	N/A	N/A	Quality control flag for CTDTEMP
CTDSAL_PSS78	N/A	3	Salinity calculated from conductivity recorded with CTD using the equation of the Practical Salinity Scale of 1978. CTDSAL_PSS78 is unitless.
CTDSAL_flag	N/A	N/A	Quality control flag for CTDSAL
Salinity_PSS78	N/A	3	Salinity calculated from conductivity measured from discrete bottles using the equation of the Practical Salinity Scale of 1978. Salinity_PSS78 is unitless.
Salinity_flag	N/A	N/A	Quality control flag for Salinity_PSS78
CTDOXY	umol/kg	1	Dissolved oxygen (O_2) content from oxygen sensors mounted on the CTD
CTDOXY_flag	N/A	N/A	Quality control flag for CTDOXY

Oxygen	umol/kg	1	Dissolved oxygen (O_2) content measured from discrete-bottle-based Winkler titration
Oxygen_flag	N/A	N/A	Quality control flag for Oxygen
DIC	umol/kg	1	Total dissolved inorganic carbon content
DIC_flag	N/A	N/A	Quality control flag for DIC
TA	umol/kg	1	Total alkalinity content
TA_flag	N/A	N/A	Quality control flag for TA
pH_T_measured	N/A	4	pH measured on Total Scale (T) at measurement temperature and 1 atmosphere pressure (0 dbar applied pressure) using spectrophotometric methods. If the pH is measured on the seawater, free, or NBS scale, replace "T" with SWS, F, or NBS, respectively. For pH measurements made using electrodes, "pH_T_measured (electrode)" should be used instead.
TEMP_pH	deg_C	2	Temperature at which the pH_TS_measured value is measured
pH_flag	N/A	N/A	Quality control flag for pH_TS_measured
Carbonated_measured	umol/kg	1	Dissolved carbonate ion content ($[CO_3^{2-}]$) at measurement temperature and 1 atmosphere pressure (0 dbar applied pressure).
TEMP_Carbonate	deg_C	2	Temperature at which the Carbonate_measured value is measured
Carbonate_flag	N/A	N/A	Quality control flag for Carbonate_measured
fCO2_measured	uatm	1	Fugacity of carbon dioxide (fCO_2) in air that is in equilibrium with seawater measured from discrete bottles at measurement temperature and 1 atmosphere pressure (0 dbar applied pressure).
TEMP_fCO2	deg_C	2	Temperature at which the *f*CO2_measured value is measured
fCO2_flag	N/A	N/A	Quality control flag for *f*CO2_measured

Silicate	umol/kg	2	Silicate (total dissolved inorganic silicate: $Si(OH)_4$, H_4SiO_4, SiO2, SiI) content
Silicate_flag	N/A	N/A	Quality control flag for Silicate
Phosphate	umol/kg	2	Phosphate (total dissolved inorganic phosphate: $H_2PO_4^-$, HPO_4^{2-}, PO_4^{3-}) content
Phosphate_flag	N/A	N/A	Quality control flag for Phosphate
Nitrate	umol/kg	2	Nitrate (NO_3^{-1}) content. This term should not be used to indicate nitrate plus nitrite content, although the distinction is generally small because nitrate >> nitrite.
Nitrate_flag	N/A	N/A	Quality control flag for Nitrate
Nitrite	umol/kg	2	Nitrite (NO_2^{-1}) content
Nitrite_flag	N/A	N/A	Quality control flag for Nitrite
Nitrate_and_Nitrite	umol/kg	2	Nitrate plus nitrite content
Nitrate_and_Nitrite_flag	N/A	N/A	Quality control flag for Nitrate_and_Nitrite
Ammonium	umol/kg	2	Ammonium (NH_4^+ and NH_3) content
Ammonium_flag	N/A	N/A	Quality control flag for Ammonium

Appendix B: Column header names for autonomous data:

OCADS Column Header standards for autonomous sensor data. Column header abbreviations, their preferred units and brief descriptions for continuous measurements of oceanographic variables using autonomous or remotely operated platforms, e.g., time-series mooring, Saildrones, gliders, and Argo floats, instead of traditional manned research vessels. "N/A" is short for "not applicable". Users can go to the "Resources" tab and click the "Quality Control Flags" collapsible to access the definitions of the quality control flags.

Column Header Standards for Autonomous Sensor Data		
Abbreviation	*Unit*	*Description*
EXPOCODE	N/A	Expedition code (EXPOCODE) consists of the four-digit International Council for the Exploration of the Sea (ICES) platform code

		and the date of departure from port (UTC) in YYYYMMDD.
Platform_type	N/A	Controlled vocabularies for the types of the platform, e.g., research vessel, Saildrone, glider, argo, fish vessel, oil tanker, mooring, etc.
Platform_name	N/A	The name of the platform, e.g., Ronald H. Brown.
Year_UTC	N/A	Calendar year in Coordinated Universal Time (UTC)
Month_UTC	N/A	Calendar month in Coordinated Universal Time (UTC)
Day_UTC	N/A	Calendar day in Coordinated Universal Time (UTC)
Time_UTC	[hh::mm::ss]	Time in the format of hh:mm:ss
Yearday_UTC	N/A	Yearday refers to the day number in an annual cycle. (e.g., 12 pm on Jan 1 means Yearday = 1.50, 6 am on Dec 31 means Yearday = 366.25 in a leap year). Two digits after the decimal point are recommended.
Latitude	decimal degree	Latitude in decimal degrees North (negative for Southern Hemisphere)
Longitude	decimal degree	Longitude in decimal degrees East (negative for Western Hemisphere)
Depth	meter	Depth (in meters) at which the sensor is located
SST_ITS90	degrees Celsius	Sea surface temperature
SSS_PSS78	N/A	Sea surface salinity
Pressure_ATM_LICOR	hPa	Atmospheric pressure as recorded by LICOR
Temperature_LICOR_ITS90	degrees Celsius	Temperature as recorded by LICOR
***x*CO2_SW_wet**	μmol/mol	Mole fraction of carbon dioxide in seawater in wet gas
***x*CO2_SW_flag**	N/A	Quality control flags for mole fraction of carbon dioxide in seawater in wet gas
***x*CO2_ATM_wet**	μmol/mol	Mole fraction of carbon dioxide in air in wet gas
***x*CO2_ATM_flag**	N/A	Quality control flags for mole fraction of carbon dioxide in air in wet gas

***x*H2O_SW**	µmol/mol	Mole fraction of H_2O in the headspace of the equilibrator
***x*H2O_ATM**	µmol/mol	Mole fraction of H_2O in air
***x*CO2_SW_dry**	µmol/mol	Mole fraction of CO_2 in seawater in dry gas
***x*CO2_ATM_dry**	µmol/mol	Mole fraction of CO_2 in air in dry gas
***f*CO2_SW_sat**	µatm	Fugacity of CO_2 in seawater at saturated water vapor pressure
***f*CO2_ATM_sat**	µatm	Fugacity of CO_2 in air at saturated water vapor pressure
d*f*CO2	µatm	Difference of fCO_2 in water and air (*f*CO2_SW - *f*CO2_Air)
pH_T_insitu	N/A	pH on total scale at in situ temperature
pH_flag	N/A	Quality control flag for pH_T_insitu
Oxygen	µmol/kg	Dissolved oxygen measured from sensor
Oxygen_flag	N/A	Quality control flag for dissolved oxygen
Percent_O2	N/A	Percent O_2 measurement made in equilibrated air
Percent_O2_flag	N/A	Quality control flag for Percent_O2
Chl_a	µg/L	Chlorophyll a concentration
Chl_a_flag	N/A	Quality control flag for Chlorophyll a concentration

Appendix C: Column header names for underway data:

OCADS Column Header standards for underway data. Column header abbreviations, their preferred units, and brief descriptions for surface underway based chemical oceanographic data. "N/A" is short for "not applicable". Users can go to the "Resources" tab and click the "Quality Control Flags" collapsible to access the definitions of the quality control flags.

Column Header Standards for Underway Data		
Abbreviation	*Unit*	*Description*
EXPOCODE	N/A	Expedition code (EXPOCODE) consists of the four-digit International Council for the Exploration of

		the Sea (ICES) platform code and the date of departure from port (UTC) in YYYYMMDD.
Cruise_ID	N/A	Cruise identification
Year_UTC	N/A	Calendar year in Coordinated Universal Time (UTC)
Month_UTC	N/A	Calendar month in Coordinated Universal Time (UTC)
Day_UTC	N/A	Calendar day in Coordinated Universal Time (UTC)
Time_UTC	[hh:mm:ss]	Time in the format of hh:mm:ss
Yearday_UTC	N/A	Yearday refers to the day number in an annual cycle. (e.g., 12 pm on Jan 1 means Yearday = 1.50, 6 am on Dec 31 means Yearday = 366.25 in a leap year). Two digits after the decimal point are recommended.
Latitude	decimal degree	Latitude in decimal degrees North (negative for Southern Hemisphere)
Longitude	decimal degree	Longitude in decimal degrees East (negative for Western Hemisphere)
Depth	m	Depth in meters
SST_ITS90	degree Celsius	Sea surface temperature (usually intake temperature)
SSS_PSS78	N/A	Sea surface salinity
Pressure_ATM	hPa	Sea level atmospheric Pressure
Temperature_EQU_ITS90	degree Celsius	Water temperature recorded in the equilibrator
Pressure_EQU	hPa	Pressure inside the headspace of the equilibrator. 1 hPa = 1 mbar.
***x*CO2_EQU**	μmol/mol	Mole fraction of carbon dioxide (dry) inside the headspace of the equilibrator
***x*CO2_ATM**	μmol/mol	Mole fraction of carbon dioxide (dry) in the atmosphere
***x*CO2_ATM_interpolated**	μmol/mol	Interpolated atmospheric xCO_2 to match with water analyses time
***f*CO2_SW_SST**	μatm	Fugacity of seawater carbon dioxide at SST
***f*CO2_SW_flag**	N/A	Quality control flag for fugacity of seawater carbon dioxide
***f*CO2_ATM_interpolated**	μatm	Interpolated atmospheric fCO_2
d*f*CO2	μatm	Delta fCO_2 between seawater and atmosphere (*f*CO2_SW - *f*CO2_ATM)

pH_T_insitu	N/A	pH on total scale at in situ temperature
pH_flag	N/A	Quality control flag for pH_T_insitu
Oxygen	µmol/kg	Dissolved oxygen measured from sensor
Oxygen_flag	N/A	Quality control flag for dissolved oxygen
Percent_O2	N/A	Percent O_2 measurement made in equilibrated air
Percent_O2_flag	N/A	Quality control flag for Percent_O2
Chl_a	µg/L	Chlorophyll a concentration
Chl_a_flag	N/A	Quality control flag Chlorophyll a concentration

Appendix D: Physiological response data:

OCADS Column Header standards for physiological data. Column header abbreviations, their preferred units, and brief descriptions for physiological response studies (e.g., laboratory experiment, mesocosm, field experiments, and natural analogues). "N/A" is short for "not applicable".

Physiological Header Descriptions		
Abbreviation	*Unit*	*Description*
Measurement_ID	N/A	A unique identification number of the measurement
Type_of_study	N/A	The specific type of physiological response study. It can be laboratory experiment, mesocosm, field experiment, natural analogues, etc. Please be as specific as possible.
Treatment_type	N/A	Whether this is an ocean acidification study or an ocean alkalinity enhancement study. The input can be a choice of "Acidification", or "Alkalinization".
Treatment_method	N/A	How specifically the seawater acidity or alkalinity was manipulated. The input can be any of "Bubbling CO_2", "Adding hydrochloric acid", "Adding calcite", "adding magnesite", or "Electrochemical", etc.
Treatment_details	N/A	The type, characteristics, and amount of alkaline or acid materials, e.g., addition of 100g of

		hydrated lime ($Ca(OH)_2$) with a grain size <63μm in diameter to a tank with a capacity of 100L
Biological_subject	N/A	Species of the marine organisms studied
Species_identification_code	N/A	It is recommended to use the species reference databases from the Integrated Taxonomic Information System (ITIS, http://www.itis.gov/), World Register of Marine Species (WoRMS, http://marinespecies.org/), Catalogue of Life (COL, https://www.catalogueoflife.org/), or Paleobiology Database (PBDB, https://paleobiodb.org/classic/classificationForm).
Life_stage	N/A	This can be egg, embryo, larva, juvenile, or adult, etc.
Location_biological_subject_collection	N/A	Location where the organisms were collected. Please be as specific as possible
Northernmost_latitude	decimal degrees	Bounding box for the location of the organism collection
Southernmost_latitude	decimal degrees	Bounding box for the location of the organism collection
Westernmost_longitude	decimal degree	Bounding box for the location of the organism collection
Easternmost_longitude	decimal degree	Bounding box for the location of the organism collection
Date_biological_subject_collection	YYYY-MM-DD	Local date when the organisms were collected
Time_biological_subject_collection	hh:mm	Local time when the organisms were collected
Experiment_location	N/A	Location where the experiment was carried out
Tank_ID	N/A	Identification of the tank (if applicable)
Tank_type	N/A	The type of the tank.
Tank_volume	L	The volume of the tank
Natural_or_artificial_seawater	N/A	Whether natural seawater or artificial seawater was used
Location_seawater_collection	N/A	If natural seawater was used, please specify the location where seawater was collected
Flow-through_or_static	N/A	Whether a flow-through system or a static system was used

Flow_rate	L/min	If a flow-through system was used, please specify the flow rate here
Target_treatment_level_*f*CO2	µatm	Target treatment level in terms of fugacity of carbon dioxide. If pCO_2 is used, please change the header name to reflect that.
Target_treatment_level_pHT	N/A	Target treatment level in terms of pH on total scale.
Target_treatment_level_TA	µmol/kg	Target treatment level in terms of total alkalinity content.
Exposure_type	N/A	Type of exposure
Date_experiment_start	YYYY-MM-DD	Local date when the experiment was started
Time_experiment_start	hh:mm	Local time when the experiment was started
Date_sampling	YYYY-MM-DD	Local date when the sample was collected
Time_sampling	hh:mm	Local time when the sample was collected
Exposure_duration	days	Duration of the exposure
Temperature_ITS90	degrees Celsius	Water temperature at ITS90 scale
Salinity_PSS78	N/A	Water salinity at the Practical Salinity Scale 1978
DIC	µmol/kg	Total dissolved inorganic carbon content
TA	µmol/kg	Total alkalinity content
pHT	N/A	pH on total scale
***f*CO2**	µatm	Fugacity of carbon dioxide. If partial pressure of carbon dioxide is reported instead, please change the header name to reflect that.
Omega_aragonite	none	Aragonite saturation state
Omega_calcite	none	Calcite saturation state
Oxygen	µmol/kg	Dissolved oxygen content
Silicate	µmol/kg	Silicate content
Phosphate	µmol/kg	Phosphate content
Nitrate	µmol/kg	Nitrate content

(Response variables)	variable dependent	

729