# Progress Report
# NOAA support for the CCHDO (CLIVAR and Carbon Hydrographic Data Office) at UCSD/SIO

Period of Activity: 01 October 2022 – 30 September 2023


**Principal Investigator**
Karen I. Stocks
Dir/Geological Data Center
UCSD Scripps Institution of Oceanography
9500 Gilman Dr., MC-0220
La Jolla, CA 92093-0220
Email: kstocks@ucsd.edu
Tel: 858-534-1898

**Financial Contact**
Rose Madson
Chief Admin Officer, GRD
UCSD Scripps Institution of Oceanography
9500 Gilman Dr., MC-0225
La Jolla, CA 92093-0225
Email: rmadson@ucsd.edu
Tel: (858) 534-4552


Rose Madson

2024-01-03 _____ 2024-01-04

Signature Date Signature Date


**Co-Principal Investigator**
Dr. Sarah Purkey
Associate Professor
UCSD Scripps Institution of Oceanography
9500 Gilman Dr., MC-0230
La Jolla, CA 92093-0230


**Budget Summary**
FY 2023: $170,00

# NOAA support for the CCHDO (CLIVAR and Carbon Hydrographic Data Office) at UCSD/SIO

Karen Stocks, Sarah Purkey
Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA

## Table of Contents



## 1. Project Summary

The CLIVAR and Carbon Hydrographic Data Office (CCHDO) is the repository, assembly center, and distribution center for global ocean CTD, hydrographic, carbon, and tracer data of the highest quality and utility. CCHDO adds value to the global observing system by ensuring that these data are curated, preserved, well-described, findable, accessible, and usable. This reduces the time to science and enlarges the user base for these data by creating standard, well-described data products.

The CCHDO currently curates data from over 2,500 cruises from 29 countries, an investment on the order of $1B in ship time. It is the official data center for CTD and water sample profile data from the Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP), a key component of both the Global Climate Observing System and the Global Ocean Observing System.

The coast-to-coast, top-to-bottom, reference-quality ocean transects curated by the CCHDO provide a keystone in assessing and understanding the extent and nature of ocean changes. They are used to monitor the ocean storage and transport of heat, freshwater, carbon and nutrients; assess the ocean's uptake of anthropogenic carbon and heat; monitor changes in ocean oxygen and acidification; and evaluate ocean circulation and ventilation. A principal scientific use of CCHDO data is model initiation, calibration and validation. Further, it provides critical reference data for Argo float calibration and quality control.

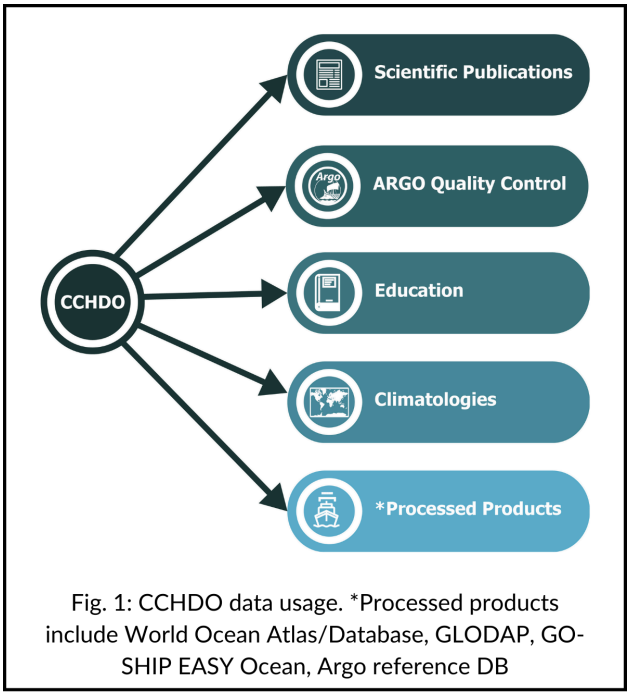


Fig. 1: CCHDO data usage. *Processed products include World Ocean Atlas/Database, GLODAP, GO-SHIP EASY Ocean, Argo reference DB

The CCHDO ensures that these valuable data and their associated documentation are (i) readily available for immediate research and education, (ii) easily accessible through several search and download sites aimed at diverse audiences, (iii) fed into global climatologies such as the World Ocean Atlas, GLODAP, GO-SHIP easy ocean, and the Argo reference climatology, (iv) supplied for climate assessments; and (v) in partnership with NOAA, archived for long-term preservation. Beyond simple accessibility, the CCHDO undertakes meticulous verification, assembly, and documentation, to produce standardized, well-described, reference-quality cruise data products from heterogeneous data submissions. The goal is to maximize data useability for the scientific community and ensure a long service life.

The CCHDO is co-funded by the National Science Foundation and NOAA; NOAA funding contributes to the overall operations of the CCHDO as well as supporting:

- serving and assembling data from cruises of particular interest to NOAA.
- collaboration with NCEI to ensure that CCHDO data can be harvested regularly by NCEI for deep archive and additional access.
- Argo's access to CCHDO data for use in quality control of Argo data.

## 2. Scientific and Observing System Accomplishments

**Progress on the milestones and performance measures in FY 2023 Work Plan**

The primary deliverables to NOAA are as follows:

A) **Posting CTD and bottle data onto the CCHDO website in their "as received" form, particularly for cruises of interest to NOAA or Argo**. This includes soliciting data submission by email and telephone to science party members, and inquiries and reminders to relevant institutions, organizations, and committees.

> **Performance Measure**: 90% of data submitted to CCHDO will be posted "as received" within 72 work hours.
>
> **Status**: metric exceeded. 96% of files were posted "as received" within 72 hours, with most posted on the day of receipt (data from two cruises required additional information from the submitting scientist before they could be posted online).

B) **Merging "as received" files into standard formats.** Assembling data into standard products, aka "merging", is the most time-intensive component of the CCHDO work. During assembly, data are:

(i) Received: Through a simple online submission page, the CCHDO takes any appropriate data set in any form and then matches the submitted measurement, metadata, and documentation with the correct measured parameter(s) and cruise.

(ii) Verified: Each file received by the CCHDO is examined. Any questions about units, quality control flags, or suspicious values are followed up with a personal email to the submitting party, ensuring the highest quality, most complete datasets get archived. While the CCHDO does not carry out subjective expert quality control, it does communicate with the submitting scientist when discrepancies are found.

(ii) Standardized: disparate files in heterogeneous formats are merged and reformatted into four standard, uniform community formats with standard parameter names, consistent across CCHDO holdings: WHP-Exchange, WOCE, netCDF (WOCE format)

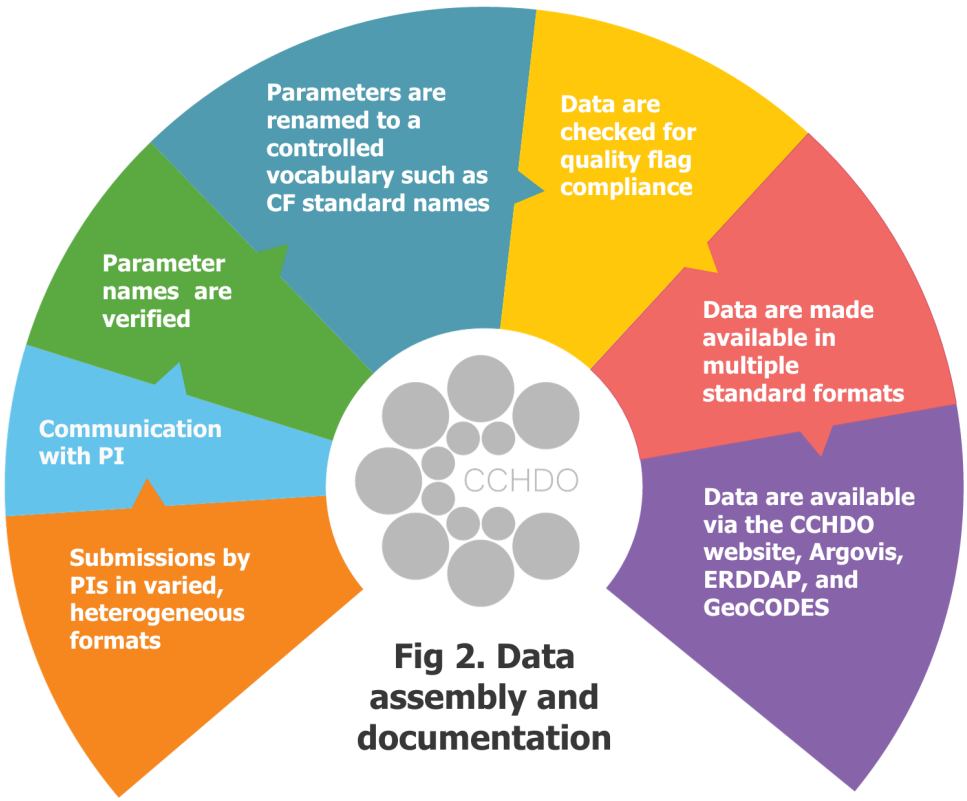


**Fig 2. Data assembly and documentation**

and CF netCDF. File-to-file near-lockstep consistency - "if you can read one CCHDO data file you can read them all" - is a key goal.

(iii) Documented and described: All documentation is co-assembled with the data, including curated, standardized, complete cruise reports. The CCHDO also tracks and documents all data modifications as well as information such as measurement and submission dates and proprietary status for each data subset.

(vi) Disseminated: through the cchdo.ucsd.edu website, each cruise has an individual webpage with standardized data products, a comprehensive data history, and "as received" files for full transparency and historical record. These are discoverable through a flexible search interface.

(vii) Archived: All data are submitted to NOAA NCEI for long-term archive, inclusion in community products such as the World Ocean Atlas, and further dissemination.

Merging a set of files into the standard products can take a few days to a few months, depending on the starting format, consistency of the data files, and responsiveness of the submitting scientists in resolving questions or issues. The resulting files are highly standardized, up-to-date, properly attributed, well-documented, and have a clear data history.

> **Performance Measure**: 18 files merged into standard data products for expeditions of interest to NOAA or Argo.
>
> **Status**: metric exceeded. 91 files were merged from 50 cruises during this reporting period (including work done under NSF co-funding).
>
> **Additional Progress**: 4 new standardized cruise reports were created and posted online, and 5 submitted, posted online as received, and in the queue for standardization.

C) **Continuing support for CCHDO data exports to NCEI and Argo**. Coordinating with NCEI and other collaborators on hydrographic data curation.

The CCHDO supports regular data downloads by NCEI (on NCEI's schedule). Argo downloads data through the main CCHDO access points and does not need a custom process, except for private data that the CCHDO pushes to a shared folder for Argo access.

D) **Technology upgrades**, including increasing cyberinfrastructure capabilities, cyber security, and the web interface

> **Data License Policy and DOIs.** DOIs and a clear data license are expected by the FAIR principles, Core Trust Seal, the Federal "Desirable Characteristics of Data Repositories

for Federally Funded Research" and by an increasing number of journals. During this reporting period, the CCHDO finalized a Data License Policy working iteratively with the GO-SHIP Data Management Team and the US GO-SHIP PI group to balance reproducibility, credit, citability, and tractability concerns: https://cchdo.ucsd.edu/policy#data_license_policy. In summary, the overall collection of data within the CCHDO is under the Creative Commons "CC 0" license (https://creativecommons.org/publicdomain/zero/1.0/) indicating free and unrestricted usage. However, upon request, the CCHDO can support a "CC BY" license which requires attribution, for specific cruises upon request.

The CCHDO then partnered with the UCSD Library Digital Collections for minting DOIs. Monthly, CCHDO exports a snapshot of the current data holdings to the Digital Collection, which receives a DOI and permanent archive (https://library.ucsd.edu/dc/collection/bb5574768j). Though citation is not legally *required* under the CC 0 license, users of data in CCHDO are *requested* to cite the DOI of the snapshot they used in data publications. The snapshot DOIs have a common stem, which will allow automated searching for citations in publication indices such as Google Scholar.

The CCHDO also initiated work to refresh two technical components of the CCHDO system. First, the "hydrotable," a tool used to track the submission of data at the parameter group level across US GO-SHIP cruises, is being ported out of a stand-alone system to an Airtable integrated with the main CCHDO cyberinfrastructure. This improves maintainability and reduces time-consuming duplicate manual data entry. This work was completed after the current reporting period ended, and has been adopted by the US GO-SHIP website.

Second, the CCHDO front-end is being moved from an older, difficult-to-edit framework into a NextJS/React framework, for maintainability, security, and flexibility. This work is expected to be completed in the coming project year.

**FY2022 Supplement Update**

In FY2022, a supplement was awarded to increase the availability of high-priority, reference-quality CTD data for Argo calibration and quality control, as well as for access by the larger scientific community. The majority of deliverables were completed within the FY2022 year however the deliverable of "A draft recommended best practices document for ship-based CTD data associated with Argo float deployments" became part of a larger best practices document, led by Tamaryn Morris and Megan Scanderbeg. This increases the reach and impact of the CTD recommendations, but required a longer publication timeline. We are pleased to report that this has now been published as a GOOS-endorsed best practice in the Ocean Best Practices System - see Morris et al. 2023 in the Publications section.

**Additional notable achievements during FY 2023**

The CCHDO was awarded a continuing 5 years of NSF funding under award 2319079 "CCHDO: serving easily accessible, global, high-quality hydrographic data for research, climate monitoring, and education". NSF co-funds the CCHDO data operations, supporting ~80% of costs, with his NOAA award supporting ~20%.

**Related work not directly funded by NOAA award**

The CCHDO continues to maintain a data portal (microstructure.ucsd.edu) for 1-m binned microstructure data ("chi-pod" data) that were digitized from the literature or submitted by PIs.

**Scientific significance**

The CCHDO adds value to the global observing system by ensuring that data from repeat hydrography expeditions are curated, preserved, well-described and documented, findable, accessible, and usable. It reduces the time to science and enlarges the user base for these data by creating standard, well-described data products.

The CCHDO data are used to address broad ocean research, including sea surface temperature and calculated surface currents, ocean heat content and transport, fluxes of heat, momentum, and freshwater and ocean carbon content and uptake. Because many CCHDO data files are of reference quality, and cover an ever-growing time span, they provide a keystone in assessing and understanding the extent and nature of ocean changes, including climate change impacts.

**Instrumental records of Essential Ocean Variables , Essential Climate Variables**,

The CCHDO does not directly collect data from instruments. Rather it serves as a repository to which sea-going programs submit their high-quality repeat CTD and bottle data. These data include multiple EOVs and ECVs.

**Issues related to funding that affect progress (e.g., reductions, delays)**

The CCHDO had two staff departures this year: long-time CCHDO member and Technical Director Steve Diggs departed to begin a position at the University Office of the President developing UC cross-campus data education and infrastructure programs. We wish him the best in this exciting role. Part-time technician Lynne Merchant also moved on to other projects. Existing staff member Andrew Barna was elevated to technical lead, and a new entry-level staff member, Andre Dos Santos was hired to support both data assembly and technical development, to fill the gaps. (These staff are shared between NSF and NOAA funding).

The one performance metric impacted by staff departures was the stand-alone presentation that Steve Diggs made annually to the Scripps graduate students - see "Outreach and Education" below.

**CCHDO Website**

Maintaining the cchdo.ucsd.edu (Fig 1) website and associated hardware and software is a primary deliverable for the CCHDO. Through the website, users can conduct quick searches based on program (e.g., GO-SHIP, SOCCOM, DIMES), ocean basin, or lat/lon box; or they can use the advanced search to combine keyword, geospatial, and temporal search criteria.

**Status:** The CCHDO website (cchdo.ucsd.edu) was operational with >99.6% uptime. The CCHDO site usage continues to be strong. Over the reporting year, the CCHDO website was viewed by 9,112 users during 17,206 sessions, from 100 countries. Website usage stats are an imperfect metric of scientific impact. They do not include access through NCEI (including derived products such as the World Ocean Atlas), ERDDAP, Argovis, or the UCSD Library Digital Collection, and exclude users who have blocked “ads” (which Google Analytic uses for tracking), leading to undercounting of science use. However, they do include “bot” crawlers that are not legitimate data users.

## 3. Outreach and Education

**Metric**: CCHDO presents to the SIO student/early career body once per year, such as at a seminar or open-house presentation.

**Status**: not met, but with significant contributions. Steve Diggs traditionally gave a guest lecture in the SIO graduate-level course Oceanographic Experiment Design and Proposal Writing. With Steve’s departure from the CCHDO, this was not done in the current project year. However, Sarah Purkey included CCHDO data examples and assignments in her teaching of SIO graduate course on Physical Oceanography, offering deeper hands-on training to a smaller selection of students. Additionally, new outreach/education activities were carried out as described below.

**Additional Progress**:

- Undergraduate student training and experience is an integral part of the CCHDO. In the past year, two students (one funded by this NOAA award, and one by the NSF award) worked with the CCHDO team. One student worked on porting the CCHDO website into a new, more flexible, and maintainable framework (in progress, expected to be completed by mid-2024), and the other worked on a natural language processing approach to extracting structured metadata from cruise reports.
- CCHDO staff participated in an in-person Early Career “Datathon” activity led by the Deep Ocean Observing Strategy Accelnet program. During this one-day activity held at Scripps, CCHDO staff gave hands-on help and training to early career researchers with data challenges in their work.
- CCHDO data are used in exercises in the textbook "Descriptive Physical Oceanography - An Introduction" by Talley, Emery, Pickard, and Swift (see http://joa.ucsd.edu/dpo).
- Presentations or posters at science conferences are used to raise awareness of the resources the CCHDO offers, and to contribute to the development and uptake of

community best practices for ocean data (These were primarily supported by non-NOAA funds):

- **Stocks KI, Aitoro E, Barna A, Berys-Gonzales C, Diggs SC, Kappa J, Purkey SB, Merchant L, Swift JH, Purkey SG** (2022) CCHDO: Your Stop for 50+ Years of Reference-Quality Hydrographic Data. AGU Fall Meeting. Chicago, 12-16 December 2022.
- Biddle M, **Berys-Gonzalez C**, **Diggs SC**, Rose K, and Muller-Karger F. ESIP and the Ocean Decade: Efforts and Opportunities to Engage. ESIP Summer Meeting. Burlington, Vermont, 18-21 July, 2023.

Technical discussions are held on the ESIP Marine Data Cluster telecons that CCHDO staff co-chair. In addition, CCHDO staff participate in committees or conferences to stay coordinated with: ARGO (Steering Team and Data Management Team), GOOS, GO-SHIP (Science Committee, US GO-SHIP PI meetings, and co-chairing the Data Management Team), NOAA-GOMO, and UN Decade of Ocean Science activities.

# 4. Publications and Reports

## *4.1. Publications by Principal Investigators*

Because the CCHDO is an operational data facility, peer-reviewed publications are not a primary output. Instead, presentations and posters at meetings are used to raise awareness of the data resources and solicit feedback.

Published as open access:

Morris, T., Scanderbeg, M., West-Mack, D., Gourcuff, C. and Poffa, N. Udaya Bhaskar, T.V.S., Hanstein, C., **Diggs, S**., Talley, L., Turpin, V., Liu, Z., Owens, B. (2023) Best practices for Core Argo floats: Getting started, physical handling, metadata, and data considerations. Version 1. Cape Town South Africa, South African Environmental Observation Network (SAEON), 45pp. DOI: https://doi.org/10.25607/OBP-1967

**Stocks, Karen**; Beaulieu, Stace; **Berys-Gonzalez, Carolina**; Biddle, Mathew; Mills, Allison; O'Connor, Sarah; et al. (2023). Easy Resources for Managing Your Ocean Data. ESIP. Figure. https://doi.org/10.6084/m9.figshare.21766709.v1 [This collaborative work was a collaboration between the ESIP Marine Data Cluster and the Deep Ocean Observing Strategy, with contributions from CCHDO staff].

## *4.2. Other Relevant Publications*

**Metric**: 18 publications added to the curated GO-SHIP list of publications annually. Note that these are not publications by project investigators, and meeting this metric is outside the direct control of the CCHDO project. Instead, they serve as an indication of the scientific impact of the CCHDO data resource.

**Status**: Not met. No publications were added since the last annual report, with a total 610 publications through time (https://www.zotero.org/groups/220378/go-ship). This online list, maintained by GO-SHIP and outside of CCHDO's control, has not been updated since April 2022 and new publications have likely come out: a Google Scholar search on publications including the term "GO-SHIP" found 222 results. These are not necessarily all publications using GO-SHIP data from the CCHDO, but they do indicate research activity not reflected in the GO-SHIP bibliography.

With respect to citation, 6475 GO-SHIP Citations have been added to the Google Scholar Repository (https://www.go-ship.org/Bib.html) for 2023, the last full reporting year (7070 have listed in 2022). We note that GO-SHIP is only one component of the CCHDO data holdings, and thus counts of GO-SHIP publications underestimate the CCHDO's impact on scientific scholarship.

## 5. Data and Publication Sharing

The CCHDO serves all its information, data, and documentation from the cchdo.ucsd.edu data portal and the UCSD Library Digital Collections (library.ucsd.edu/dc/collection/bb5574768j), and serves data through the ArgoVis portal (argovis.colorado.edu) and the NOAA NMFS ERDDAP server (data.pmel.noaa.gov/generic/erddap/tabledap/cchdo_bottle.html and data.pmel.noaa.gov/generic/erddap/tabledap/cchdo_ctd.html). It also contributes its data to the NOAA National Centers for Environmental Information for long-term preservation, and additional access, integration into global products.

The CCHDO receives a small amount of special, rapid-access CTD data for Argo calibration as part of its working relationship with the Argo project. These data are kept sequestered from public release as per agreements with the data originators, Argo, oversight bodies, and funding agencies, though is made public as soon as released.

## 6. Project Highlight Slides

See separate Powerpoint file.