

NOAA Support for the CLIVAR and Carbon Hydrographic Data Office at UCSD/SIO, 2013-2015

Dr. James H. Swift and Dr. T. Bruce Appelgate
UCSD Scripps Institution of Oceanography, La Jolla, CA

Table of Contents

1. Project Summary.....	2
2. Scientific and Observing System Accomplishments	3
3. Outreach and Education.....	7
4. Publications and Reports.....	7
4.1. Publications by Principal Investigators.....	7
4.2. Other Relevant Publications	7
5. Slides.....	8

1. Project Summary

The CLIVAR and Carbon Hydrographic Data Office (CCHDO) brings together, verifies, and corrects content and format errors in hydrographic and tracer data used in large-scale ocean carbon, global change, water mass, and circulation studies [CLIVAR = Climate Variability and Predictability, a component of the World Climate Research Program]. It assembles the data with relevant documentation, and carefully prepares them for dissemination and archive. In addition it works to promote appropriate methodology across a wide range of process studies and observing systems, applicable community standards, communications, and data compatibility. The CCHDO brings data together to a common content and readability standard, thereby greatly reducing the difficulties research and education data users encounter. Documentation associated with the data are collected, reorganized to a common standard (where possible), and preserved with the data. The CCHDO makes it possible for all data users to cope with the temporal-, content-, and format-related file diversity the different originators engender without requiring each user to hire data specialists to rework the data. The CCHDO supports CLIVAR and ocean carbon science programs, and is a critical component of a global observing system for the physical climate/CO₂ system. The CCHDO is the official DAC for the GO-SHIP program and provides value-added public and non-public CTD data used for verification of data from the international Argo and OceanSITES programs.

CCHDO data are free and available (without passwords) in the community "WHP-Exchange" format, legacy WOCE format, and netCDF format from its web site <http://cchdo.ucsd.edu>. Fully CF-compliant netCDF data files are provided to both the Argo and OceanSITES programs.

The largest share of CCHDO support is provided by NSF, with about 20% of funding presently from NOAA. CCHDO activities of special interest to NOAA, and for which NOAA support is applied, include:

1. Provide data from specific cruises that are of special interest to NOAA.

2. Continue working to improve the relationship with NOAA's National Oceanographic Data Center (NODC) via:
 - a. Continued data assembly of cruise hydrographic data and metadata, particularly those from the GO-SHIP program. [GO-SHIP = the Global Ocean Ship-Based Hydrographic Investigations Program.]
 - b. Working with NODC to improve efficiency of transfer of data and to make CCHDO data more "archive ready".
 - c. Improved CCHDO participation in discussions towards enhancing integration of specific related Data Assembly Centers to reduce ambiguity and redundancy in data archiving.
3. Broaden and simplify the accessibility of CCHDO data sets.
4. Support CCHDO Staff members' activities with related data systems and activities (Argo, OceanSITES, etc.) which are of special interest to NOAA.

2. Scientific and Observing System Accomplishments

During the reporting period the CCHDO continued to: increase its US and non-US CTD profiles, including both public and non-public data available for Argo reference; increase its data holdings of US, GO-SHIP, and many other cruise data; reconcile EXPOCODE expedition identifiers among US data centers; and improve data search capabilities and bulk data download on the CCHDO web site. An Applications Programming Interface (API) to enable large data users (e.g., modelers) to directly download any/all data of interest is in alpha testing, but the issue of direct bulk data downloads remains a thorny one due partly to the ever-changing nature of CCHDO holdings and the impacts of upgrades.

There is a substantial amount of CTD data of value to the Argo and ocean research/modeling communities. The CCHDO continues to work with NOAA to assemble and incorporate into NOAA-related holdings select cruise data sets. [For example, 1022 HOT and BATS profiles from 109 cruises were submitted to OceanSITES. Also, more than 1400 new CTD profiles from US and non-US cruises were added to the Argo reference CTD data set in the last 12 months (2013-2014) between Argo Data Management Committee meetings.]

The NOAA/NODC works with several data assembly centers, including the CCHDO, to bring data into NODC. The CCHDO continues to improve this relationship with NODC in the following areas:

- i. Continued data assembly of cruise hydrographic data and metadata, particularly those from the GO-SHIP program. [Also, the CCHDO briefly hosted the GO-SHIP web site as a transition to an improved site now being hosted at JCOMMOPS, assisted the new GO-SHIP technical manager with start-up; and held a 2-day meeting with GO-SHIP and US HYDRO leadership.]
- ii. Working with NODC to improve efficiency of transfer of data and to make CCHDO data more "archive ready". We are transitioning from a set of NODC-based Perl scripts which rely on a static URI model to a more modern API which will allow a more dynamic and robust transfer of the CCHDO data on-demand.
- iii Working with Tim Boyer (NODC) to better synchronize data holdings.

- iv. Improved CCHDO participation in discussions towards enhancing integration of the related Data Assembly Centers (specifically CDIAC and BCO-DMO), to reduce ambiguity and redundancy in data archiving. [For example, an EXPOCODE reconciliation has taken place between the CCHDO, BCO-DMO, and CDIAC (this also reconciles GO-SHIP EXPOCODES since the CCHDO is the data assembly center for GO-SHIP.)]

CCHDO holdings are of value to a wide audience of climate researchers (including modelers) and other users. The CCHDO has implemented changes to make their holdings more query-able and accessible through modern data management standards and practices (including those that NOAA and the international oceanographic community recommend). For example, significant progress has been made regarding data searches based on available parameters, which will soon be able to be combined with other search features, and the "OR" function will soon be available in data searches. A data cart now enables bulk download of selected files.

A broad international community makes use of CCHDO data, with use most intense from US researchers and students. For example, during the period July 1, 2012 - July 1, 2013 these are the visit statistics for the CCHDO web site <http://cchdo.ucsd.edu> (counting is done by IP address):

- 12,091 visits,
- 37 countries with visitors who interacted with the site,
- 4,722 users,
- 7,620 visits from repeat users, and
- 4,471 visits from new users.

[To eliminate in-house use, the above does not count visits originating from within UCSD. This does, however, also eliminate inclusion of use of the CCHDO website by SIO researchers and students, which is significant.]

Data submissions and subsequent CCHDO web site data updates continue to be frequent. For example, these are data submission statistics for the period Oct 1, 2013 - Sep 30, 2014 :

	Files	Cruises
SUBMITTED Files (new and updated)	101	54
WEBSITE UPDATES		
Bottle	105	69
CTD	55	46
SUM	64	42
files on line but not yet checked	126	

Over that same period the CCHDO added 3567 new pages of documentation to the website. These pages represent 55 unique cruises. And during this time period the data histories were updated 859 times as part of 212 data file submissions and 647 website updates.

In terms of data submissions, participation from most US data originators (the PIs who are in charge of one or more parameters) has been excellent, with only one US PI (not NOAA funded) overdue with data deliveries, and recent progress has been made in obtaining all of those data.

Regarding international participation, there are seven nations engaged in full-depth repeats of WHP basin-scale transects which include ocean carbon parameters: Australia, Canada, France, Germany, Japan, the UK, and the US. (The CCHDO also hosts data with less broad parameter content from other nations.) Of these, Australian, Japanese, and US investigators continue to routinely provide their CTD and bottle data to the CCHDO according to each nation's schedule, and data receipts from Canada from the Bedford Institute of Oceanography have become routine on the schedule they set. Some UK investigators by-pass their national data center to provide CTD/hydrographic data to the CCHDO on the schedule they set (sometimes preliminary shipboard data are sent immediately after a cruise). Through attendance at meetings (OceanSITES, Argo, DIMES, etc.), and the CDIAC and Dr. Robert Key (Princeton), the CCHDO receives an increasing stream of data from other countries. Because the CCHDO is now more widely recognized as the CTD/hydrographic/tracer/carbon data center for international GO-SHIP, data receipts from the GO-SHIP community are now more nearly straightforward, prompt, and complete.

International coordination continues to improve, partly due to CCHDO efforts and partly due to the growing success of the international scientific oversight and planning body GO-SHIP, which brings together scientists with interests in physical oceanography, the carbon cycle, marine biogeochemistry and ecosystems, and other users and collectors of hydrographic data to develop a globally coordinated network of sustained hydrographic sections as part of the global ocean/climate observing system. GO-SHIP has become a widely known and appreciated effort within the community. The relationship with GO-SHIP assists the CCHDO in keeping up to date with international cruises of interest to the CCHDO's data users.

The CCHDO continues to enjoy a mutually beneficial relationship with both the Argo and OceanSITES programs. In exchange for the CCHDO providing both programs with specifically formatted CTD data, these NOAA programs in turn assist the CCHDO in the discovery and acquisition of hydrographic data from the PIs involved in those communities. CCHDO data management support of the NSF-funded DIMES project has developed well, especially regarding data access and distribution. The tools, formats and methods that NSF and NOAA have invested in for the CCHDO have proven to be directly applicable to effective management of data for these types of process studies and field programs.

Since the time of WOCE and the WHPO (the CCHDO's predecessor), the CCHDO has provided direct benefits to NOAA and NOAA researchers, including, for example, supplying a full range of data and documentation services for NOAA cruises for the program, providing data to many NOAA scientists who routinely use CCHDO data and documentation, supporting CTD data services for the NOAA-supported Ocean-SITES program, and providing significant "front-end" data services which aid NOAA/NODC. The CCHDO's present, partial fiscal support from NOAA helps to support the above activities and to provide data from specific cruises that are of special interest to NOAA, to work more closely with NOAA on data assembly as related to NOAA data centers, to broaden and simplify the accessibility of CCHDO data sets to NOAA

investigators, and to make holdings more query-able and accessible through modern data management standards and practices (including those that NOAA and the international oceanographic community recommend).

The CCHDO maintains files for more than 1300 cruises. The underpinnings of the CCHDO's in-house (staff only) and on-line (public) operations require continual modifications and updates to maintain compliance with security and operational guidelines, always with a focus on ease and power of use combined with underlying simplicity of operations and maintenance. The CCHDO continues to expand automation of routine tasks to speed the work of the data specialists and reduce time, errors, inconsistencies, and omissions.

The CCHDO has (internal) software modules that read and write data in WOCE, WHP-Exchange, netCDF, OceanSITES, MATLAB, and some "in-house" formats used by major data submitters. This software includes some unit conversions, conversion of depths to pressures, reorder of rows based on pressure, and so forth. A module has been developed to read some SeaBird CTD files directly. The new and updated modules resulted in large performance increases for reading and writing data formats.

The goal and result of CCHDO activities is an ever-growing, publicly-available library of World Ocean CTD, hydrographic, ocean carbon, and tracer data which are correct, up-to-date, properly attributed, well-documented, and with a clear data history. The CCHDO is a stable, mature operation with a consistent goal: supplying a broad community with a dependable data set meeting community needs. The CCHDO's multi-year strategy for utilizing its NSF and NOAA support includes continuing to broaden its reach to more of the data originators who generate the data the community wishes to obtain from the CCHDO in consistent form, and continuing to work with all data providers to assist their transfer of data and documentation to the CCHDO in forms that mutually reduce the workload on them and the CCHDO. At the same time, internal CCHDO operations are continually being examined and improved for greater reliability, accuracy, and efficiency. Thus, an increasing volume of data can be handled from an increasing number of data originators, and supplied to an increasing number of data users, with minimal changes in CCHDO staff or inflation/merit-adjusted agency support.

There are limits to efficiency improvements. Simply put, the data handled by the CCHDO are products of human endeavor and so contain errors and inconsistencies which require evaluation by data specialists. A data specialist working for a reference data center such as the CCHDO cannot blithely quick-correct or simply ignore many of the subtle or confusing issues which arise in handling data. Meeting the responsibility to "get it right" can require serious expertise, attention and time. With the ever-increasing volume of international and quick-release data, the data and documentation backlogs faced by the CCHDO's specialists are increasing. We judge that to work the data stream in the timely manner expected by the community of CCHDO data users, at a minimum the CCHDO requires funds to hire, train, and support at least one additional data specialist to handle principal matters, and to hire 1-2 more undergraduate student research assistants to carry out more nearly routine data and documentation tasks. This would require approximately \$100k more per year in agency funding, i.e. from NSF and/or NOAA. The interim measure being applied is to prioritize data tasks with US CO₂/repeat hydrography first, International GO-SHIP second, and issues for other cruises third. This does, however, ensure an

ever-growing backlog of not-yet-addressed data tasks. To ensure that users at least have access to them, all new data files not yet groomed by the CCHDO data specialists are immediately placed on-line in a special category labeled "as received" data.

The CCHDO's on-line pages are live-generated from the CCHDO database of cruise information so that the CCHDO's data users are literally as up to date as the CCHDO is itself. Most data files - especially for cruises supported by US funding agencies - contain citation information in the files themselves, and the CCHDO web site pages acknowledge support from NSF and NOAA. Data access problems reported by users are addressed immediately. The CCHDO's data holdings are regularly harvested (approximately quarterly) by NOAA/NODC for new data and updates.

The data files curated and on-line at the CCHDO are in wide use by US and international research communities working to address key questions about the state of the world's oceans and their regional variations. The CCHDO data are used to address 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 of the CCHDO data are considered to be of reference quality, the CCHDO data cover an ever-growing time span, thus providing a keystone in assessing and understanding the extent and nature of ocean changes.

3. Outreach and Education

The CCHDO maintains a website (<http://cchdo.ucsd.edu>) where the data and accomplishments of the project are maintained for public and scientific view and use. CCHDO data form the core data used in the exercises developed for the textbook "Descriptive Physical Oceanography - An Introduction" by Talley, Emery, Pickard, and Swift (see <http://joa.ucsd.edu/dpo>). The CCHDO contributed to and maintains the outreach pages on the USHYDRO web site, including a virtual cruise (ship plans, photos, videos, etc.) and an example of a complete cruise from proposal through preliminary data interpretation (see <http://ushydro.ucsd.edu/outreach>). Also, the USHYDRO site includes blogs from students participating at sea. Additional community outreach is accomplished by attendance and poster presentations at conferences and meetings.

4. Publications and Reports

4.1. Publications by Principal Investigators

The CCHDO is a data assembly and distribution center. The work plan for the NOAA funding for the CCHDO does not involve traditional research activities that lead to peer-reviewed publications from the CCHDO Principal Investigators. Also, the NOAA funding does not support preparation of peer-reviewed publications by the CCHDO PIs using NOAA OCO funds..

4.2. Other Relevant Publications

Many peer-reviewed publications are generated by PIs from a wide range of institutions and nations who use data provided by the CCHDO, including data whose assembly and distribution are funded by NOAA support of the CCHDO. There is, however, no viable mechanism to track the publications generated by the thousands of data requests during the past year. The

international GO-SHIP office last year initiated a bibliography of research publications resulting from the Global Ocean Carbon and Repeat Hydrography program, for which the data are obtained from the CCHDO. The present contents of the bibliography may be viewed at the GO-SHIP web site (<http://www.go-ship.org/Bib.html>). An interactive on-line version of a similar bibliography is available on the USHYDRO web site (<http://ushydro.ucsd.edu/bibliography/>).

5. Slides

We have attached a PowerPoint with three slides illustrating the status of our project.