OceanSITES Data Management Enhancement

Period of Activity: 01 October 2022 – 30 September 2023

Principal Investigator
Name: Eugene F. Burger
Institute: NOAA/OAR/PMEL
Address: 7600 Sand Point Way NE, Seattle WA 98155
Email: eugene.burger@noaa.gov
Tel: (206) 526-4586

Financial Contact
Name: Ogie Olanday
Institute: NOAA/OAR/PMEL
Address: 7600 Sand Point Way NE, Seattle WA 98155
Email: ogie.a.olanday@noaa.gov
Tel: (206) 526-6236

Signature Date

Signature Date

Co-Principal Investigator
Dr Meghan Cronin
NOAA/OAR/PMEL
7600 Sand Point Way NE
Seattle, WA 98115
(206) 526-526-6449
meghan.f.cronin@noaa.gov

Co-Principal Investigator
Kevin M. O'Brien
UW/CICOES - PMEL
7600 Sand Point Way NE
Seattle, WA 98115
(206) 526-526-6751
kevin.m.o'brien@noaa.gov

Co-Principal Investigator
Dr. Dongxiao Zhang
UW/CICOES - PMEL
7600 Sand Point Way NE
Seattle, WA 98115
(206) 526-526-4184
dongxiao.zhang@noaa.gov

Budget Summary
FY 2022: Received: $287,900

OceanSITES Data Management Enhancement

Eugene F. Burger[1], Dr. Meghan Cronin[2], Kevin O'Brien[3], Dr. Dongxiao Zhang[4]
[1, 2]NOAA/OAR/PMEL, Seattle, WA
[2, 3]University of Washington, CICOES/PMEL, Seattle, WA

Table of Contents

1. Project Summary

OceanSITES, as an authoritative and trusted organization, seeks to better serve time series of global ocean observations obtained by projects worldwide. While the data formats are uniform, each project has its own mission objective and time frame, and most of the OceanSITES data are not served as long aggregated time series. Though the OceanSITES community identified the long time series as a critical OceanSITES product, historically, these time series were not sent to the GDAC.

The project objective is to improve the OceanSITES data systems for both data providers and data users, and in this way make the OceanSITES data more Findable, Accessible, Interoperable and Reusable (FAIR). By doing this, the OceanSITES data holdings will be increased through increased data submissions, and OceanSITES data will be more usable and valuable. This effort

hopes to accomplish these through software development, establishing improved metadata interchange, reduced data submission overhead, and effective data discovery and access.

2. Development in support of OceanSITES Data Management

Data discovery and download
This year's development focused on data discovery and data download dashboard prototyping. The development of two dashboards developed in the previous year involved significant refactoring, while another dashboard was developed to an advanced prototype stage. The OceanSITES Long Time Series and Air-Sea fluxes dashboards are now publicly available to users.

A third thematic data visualization and data access dashboard was prototyped with input from some OceanSITES PIs. This dashboard focuses on subsurface time series. A technical challenge with the DeepTS dashboard is the volume of data served for time series that span multiple decades. An approach was developed that allows rapid data exploration, and permits the rendering of the data visualization without any data decimation.

Dashboards
Version 1 of the Long Time Series and Air-Sea Flux dashboards were presented to OceanSITES data practitioners at the Data Management Team (DMT) meetings, where it received favorable feedback. Further discussion with PI's highlighted shortcomings in the functionality, principally to visually identify platforms with data. The two dashboards were refactored and the updated dashboards were tested and deployed for public access.

To present this immediate user feedback, the data discovery web application required comprehensive metadata on data availability. These metadata should also be hosted in an interface that allows rapid querying. This metadata "product" and related infrastructure is required to ensure a responsive data discovery experience for end users.

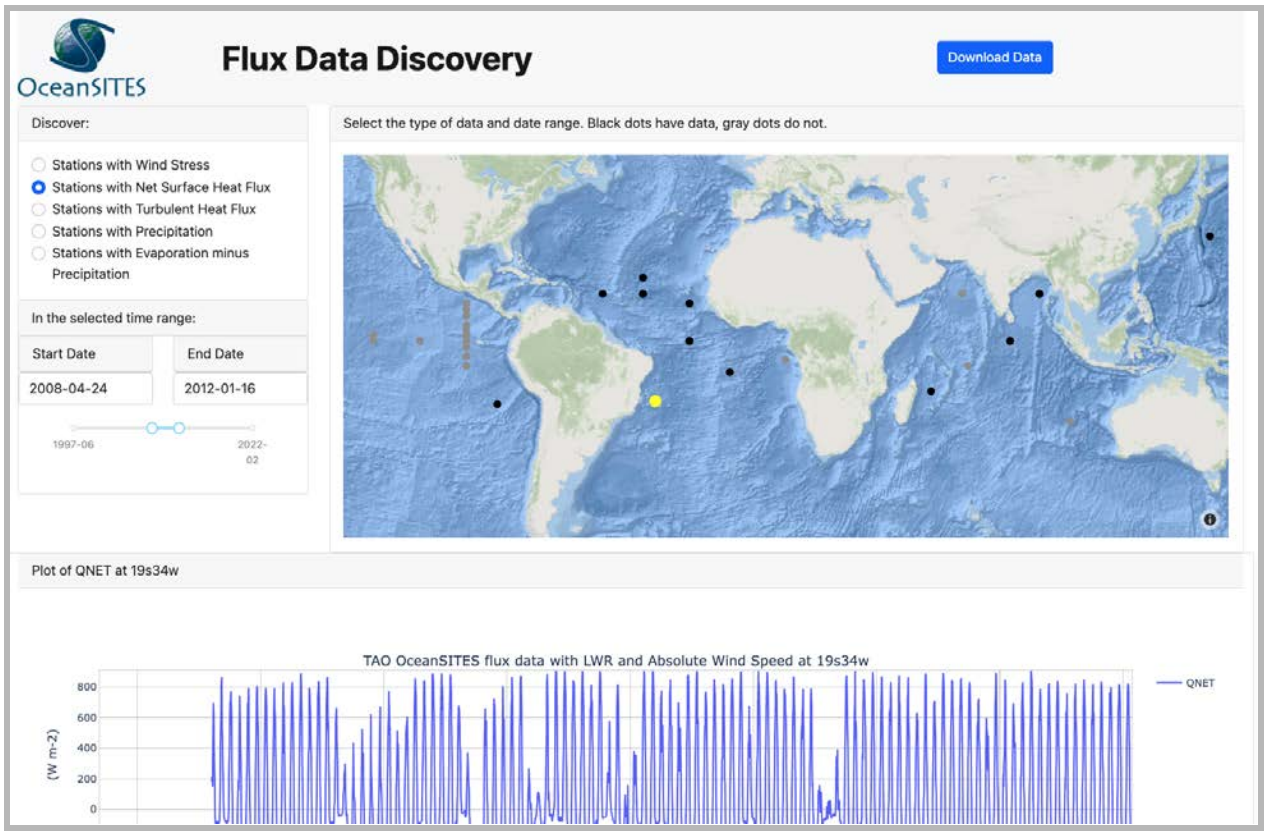

Figure 1: The revised Ocean Flux-focused data discovery and download dashboard.

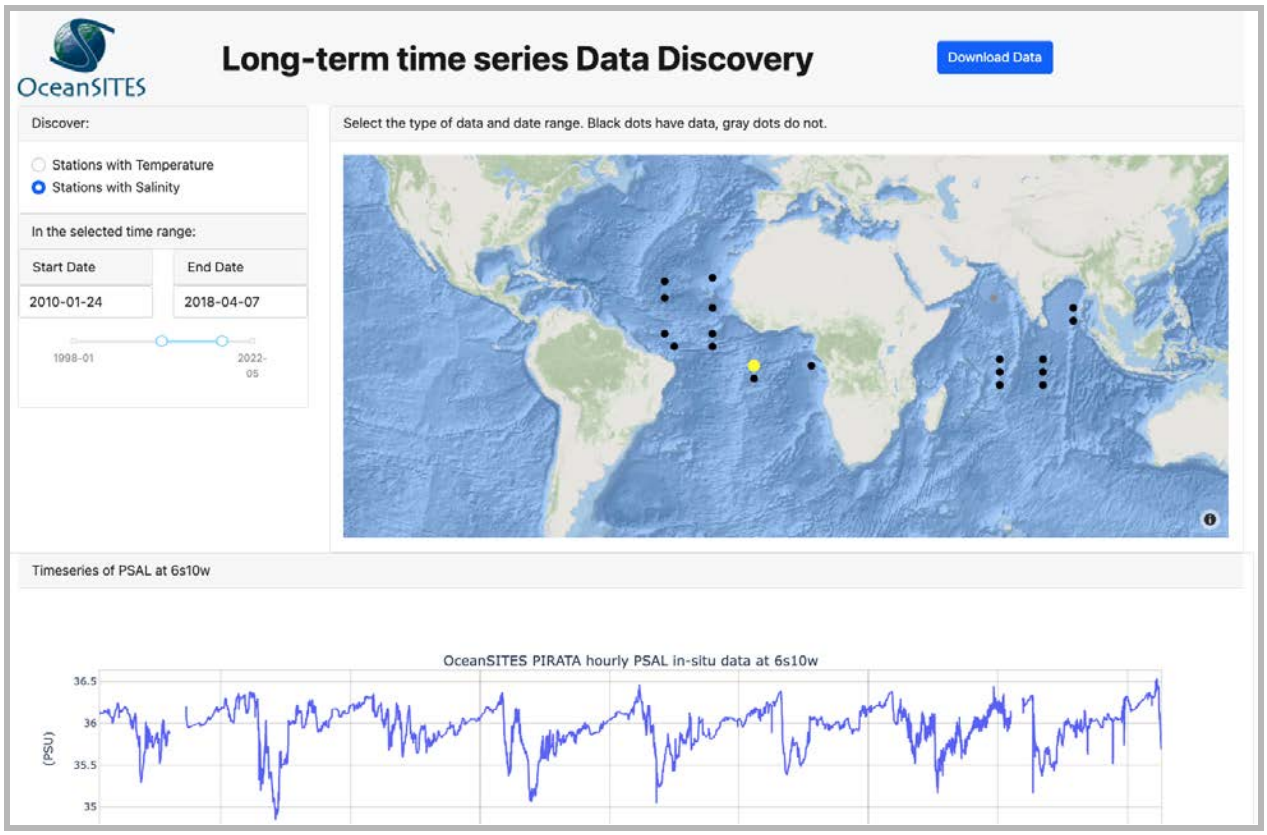

Figure 2. The revised Long-term time series focussed data discovery and download dashboard.

The Long Time Series and Air-Sea Flux dashboard used a new metadata assembly approach for rapid data discovery. Developers were presented with a new challenge to present a responsive interface for data discovery on a large volume of data.

For the Deep Time Series (DeepTS) dashboard, the same minimalist variable selection was implemented with a selection of water temperature, salinity and conductivity as selectable parameters. The variables can be used in combination with the time selector as a first data selection choice. The combination of the variable and time selection will highlight a mooring on the map that has data for the variable and time selected.

With time series at multiple depths, the data volume can be significant. The approach is to render an initial time series, appropriate for the time selected that optimizes the visualization render time. Figure 3 shows the results from such a selection.

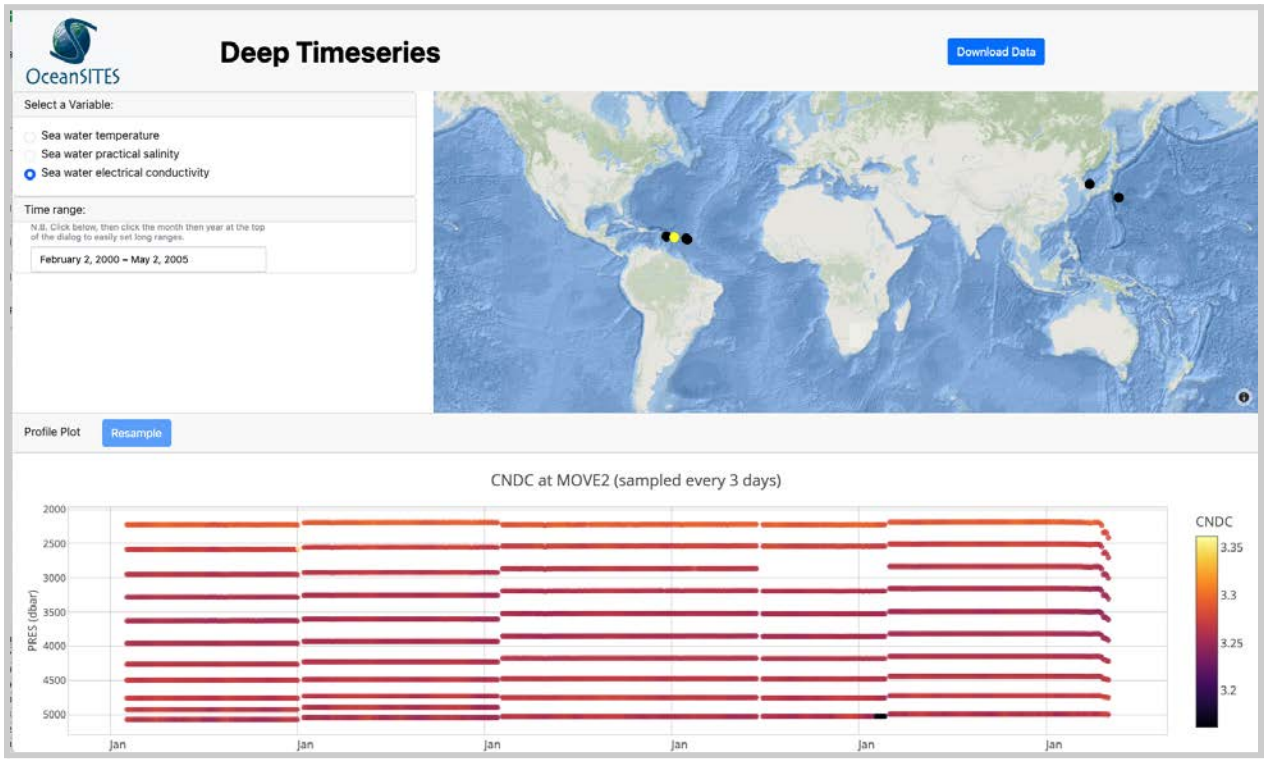

Figure 3. The Deep Timerieis prototype dashboard with results from an initial data selection.

The initial data visualization renders both a depth plot and a time series plot, as seen in Figure 4. These plots are typically rendered seconds after the initial data selection.

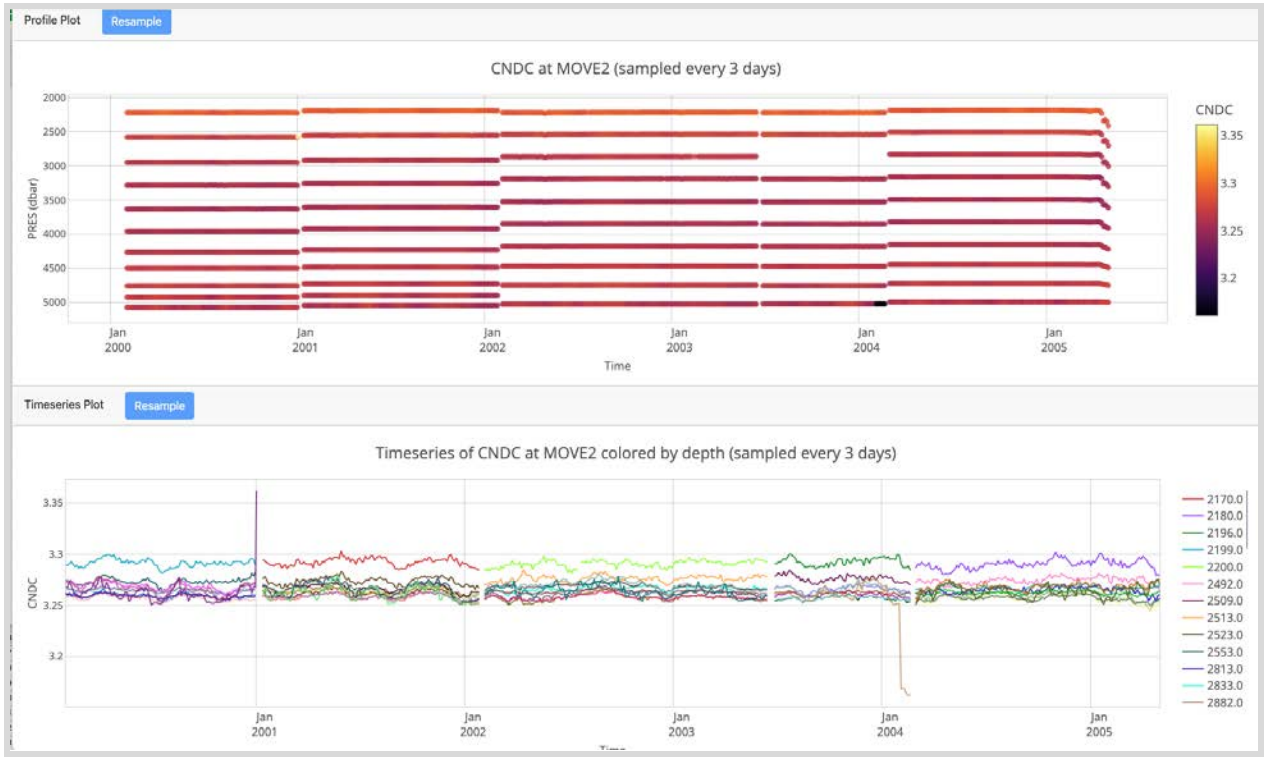

Figure 4. The colored rendered depth plot and time series plots rendered after initial data selection.

The decimated data are displayed as the user zooms in on the data. This, again, allows the user to rapidly sub-select data to identify areas of interest. Figure 5 illustrates the display when a user zooms in.

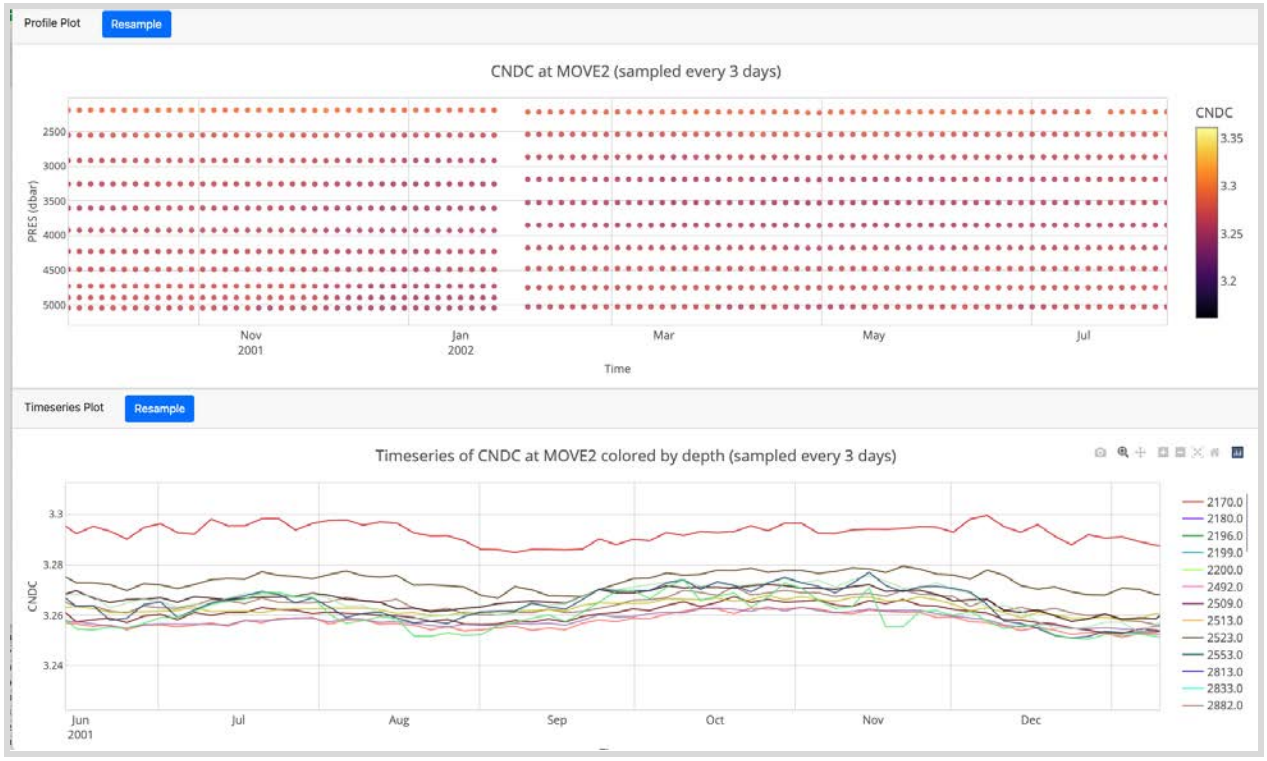

Figure 5. Data visualization rendering of decimated datasets.

For the user to get the full, non-decimated data rendering, the user clicks the *Resample* button, and the dataset is again resampled to provide an appropriate visualization. Figure 6 illustrates the same selections as in Figure 5, but with the data resampled to display the optimum data visualization for the period selected. This process can be repeated until the user has identified the variable and period of interest. The data decimation scheme allows for rapid data inspection. At any time, the user can click the download button, and the data rendered in the visualization will be downloaded.

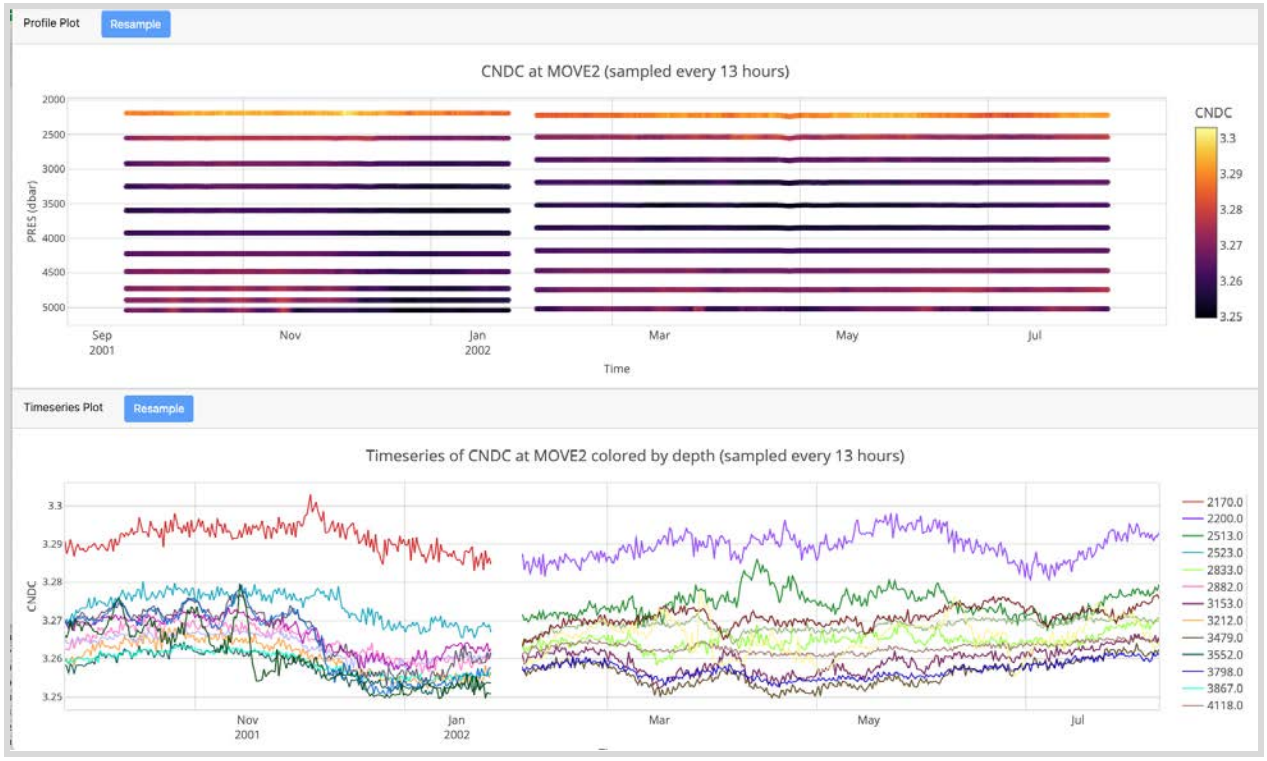

Figure 6. The data selection similar to that in Figure 5, resampled.

Once the data are selected, the user is presented with visualizations of these data. These data visualizations are fully interactive with the user's ability to zoom or manipulate the visualization (Figure 4). A local copy of the visualization for use in another document is a click away with the "snapshot" icon with each plot.

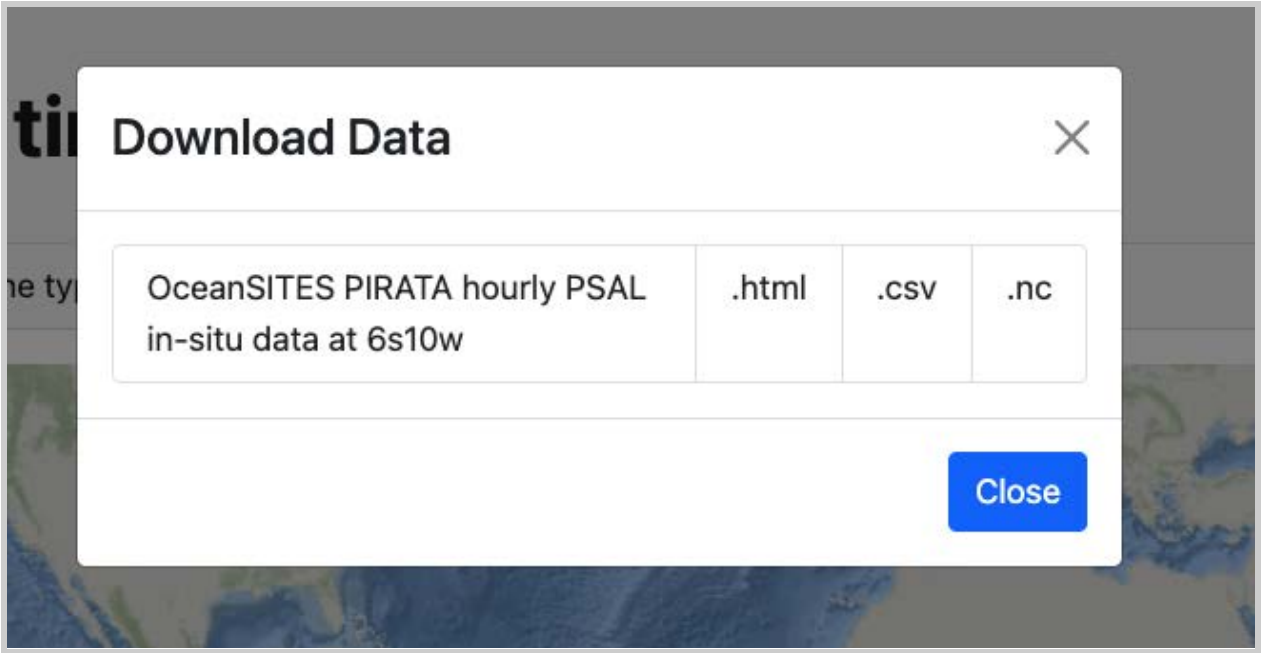

Figure 7. The data download format selection interface.

Data hosting
Data served through these sites were hosted in an ERDDAP server at PMEL, and more data from the TAO and PIRATA OceranSITES mooting were added to this PMEL hosted data collection. The emphasis is on adding the deep time series data. The objective is to move these data and the ERDDAP server to the NDBC GDAC.

Data dashboard infrastructure
To support the data dashboard environment, PMEL has, with support from this project, invested in a Plotly Dash Enterprise development and hosting environment. The licenses were delivered in June 2022, and acceptance testing has been completed. This investment will speed up further web development as deployment to a secure environment much more rapidly. An added benefit is the security environment combined with the development environment. This benefits this project as the risk of downtime due to IOT security vulnerabilities or compromise is reduced. This platform is now being upgraded to the last version (5.2) of the Dash infrastructure.

Data submission
Streamlined data submission that shifts data formatting complexity from the PI to the data submission platform is one of the pillars for this project. Core functionality will be based on the

dashboard developed for the Surface Ocean CO2 (SOCAT) product, and later modified for the Ocean Acidification Program. This dashboard will, however, be developed on the Dash Enterprise framework to allow more rapid development and deployment. As much of the functionality for data submission and metadata acceptance overlaps with requirements for the Ocean Acidification program, the development benefited from development undertaken for both OAP and OceanSITES.

PI Burger met with the NDBC OceanSITES PI, Karen Grissom and the NDBC data team. Another meeting will happen with co-PI O'Brien to discuss development plans for infrastructure hosting at NDBC.

Engagement with OceanSites Data Management Community
PI Burger and Co-PI O'Brien have been active in the OceanSITES data management endeavors through the Data Management Team (DMT).

Data curator
The position description for the data curator has been completed and this position will be listed early in calendar year 2024, following approval from the PMEL lab Director.

PI Burger continued his role as chair of the OceanSITES DMT, and has been involved in the monthly OceanSITES data Management meetings. This engagement is essential towards a better understanding of the OceanSITES data needs and meeting and improving these data and metadata management standards through the development efforts described here. Initial developments were presented at the DMT

Development Summary
This site is already publicly available for early evaluation by OceanSITES collaborators. The Deep TS dashboard will be rolled out soon.

3. Outreach and Education

PMEL developers, and co-PIs are engaged in multiple aspects of OceanSITES, most particularly the data management aspects of OceanSITES. Presentations and demonstrations have been given to both the OceanSITES DMT and OceanSITES Executive Committees.

4. Publications and Reports

FY2023 Annual Report OceanSITES Data Management Enhancement

4.1. Publications by Principal Investigators

a. No publication can be listed. PI Burger demonstrated the data visualization application to the OceanSITES Executive Team, the OceanSITES Data Management Teams and at the annual OceanSITE meeting.

4.2. Other Relevant Publications

None

5. Data and Publication Sharing

N/A
