Streamlined Submission and Archival of Quality-Checked Carbon Data

Linus Kamb[2], Eugene F. Burger[1], Roland Schweitzer[3], Kevin O'Brien[2]

[1] NOAA/PMEL, Seattle, WA; [2] University of Washington – CICOES/NOAA/PMEL, Seattle, WA; [3] Weathertop Consulting

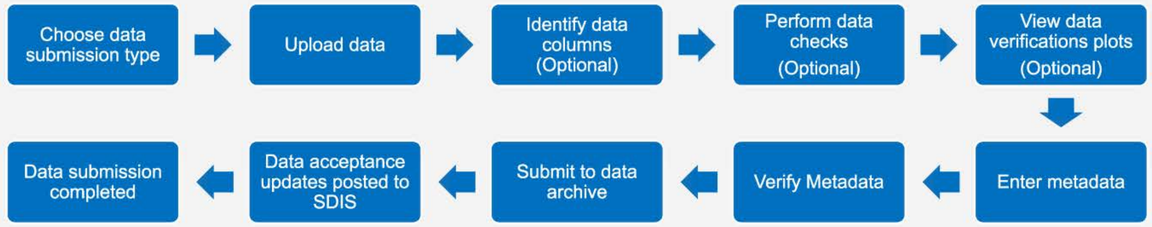
Cooperative Institute for CLIMATE, OCEAN & ECOSYSTEM STUDIES

The Scientific Data Information System

The OAP **The Scientific Data Information System (SDIS)** enables OAP funded scientists to efficiently submit their data and metadata he the NOAA Ocean Acidification Data System (OADS). This system has been operational for two years. To ensure application integrity and meet user needs, updates and new capabilities are continuously added.

The SDIS system adds value to the data during data submission with functionality to verify the data before submission, and an easy to use metadata entry interface. Use of the additional functionality is left to the end-user, with only metadata submission that is required.

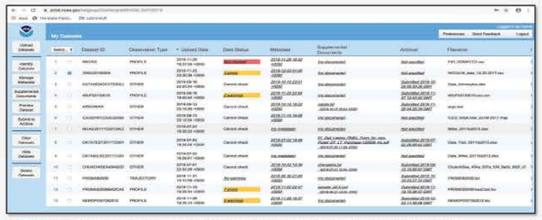
SDIS Application Workflow



1. Easy Data Upload and Data Check Highlights

Data Submission

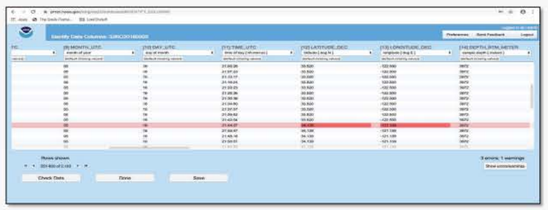
Data can be submitted in human readable and easily editable comma separated value (CSV) or Excel format. The data submission tool recognizes frequently submitted variables and identifies these in submitted data. This **allows ease and flexibility in data submission.**



User dashboard shows uploaded data and submission status

Built-in Data Sanity Check (optional)

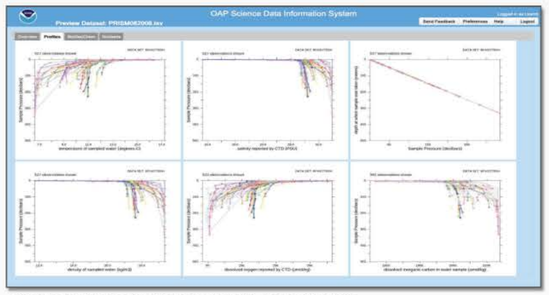
With the data properly identified, the sanity check warns the user if data are outside the bounds of pre-set data limits. Examples of data checks include out-of-bounds values, inconsistent latitude, longitude, depth values, or data submitted in an incorrect unit. Columns or individual records with errors are highlighted to indicate flagged values.



Errors and warnings detected by the Sanity Check are highlighted

Data Preview

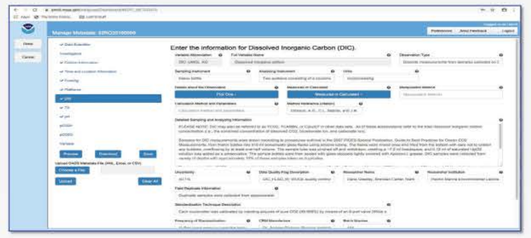
A collection of preview plots and maps allow the user to assess data integrity. A variety of plots showing overview information such as observation locations and a selection of property-property plots can highlight obvious data errors. This step **improves data quality** by reducing common data mistakes before the data can be submitted.



Optional Property-Property plots to verify data

2. Integrated Metadata Entry Tool

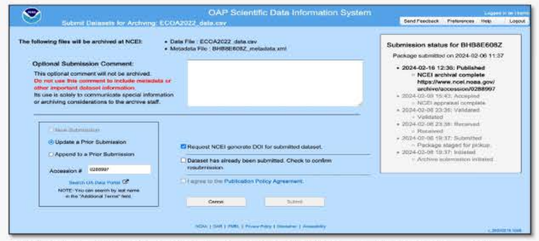
Some metadata are extracted from uploaded data. These extracted metadata are pre-populated in the metadata tool integrated into the data upload dashboard. Completed metadata as well as base reusable templates can be uploaded in Excel, CSV, or XML formats.



The metadata entry and upload tool is integrated with other components

3. Streamlined Data Archival

Archiving the high quality, high value data and metadata to a National Archive Center of choice ensures **long term preservation**. Using services, developed in collaboration with NCEI (for US submitters), data submission effort is reduced to a few button-clicks. Streamlined archival processes **reduce the overhead for scientists to meet their data management obligations**. User options will be added for the submitter to select the archive destination.



After data files are uploaded and metadata completion, the data and metadata package is submitted to the archive with a few button clicks.

High Quality Datasets, Low Data Management Burden

These tools and workflow reduces the data management burden for scientists, while at the same time delivering high quality data in interoperable and standards-based formats that promote easier discovery and use of these high-value data. These data processes will help scientists meet their obligations for data documentation, data access, and archival.

In addition, providing standardized, complete, rich metadata, and a quality checked and standardized data submission package to the Archive enables greater automation of the archival process, allowing the data to be made available more quickly.