

# Streamlining Oceanic Biogeochemical Dataset Assembly in Support of Global Data Products

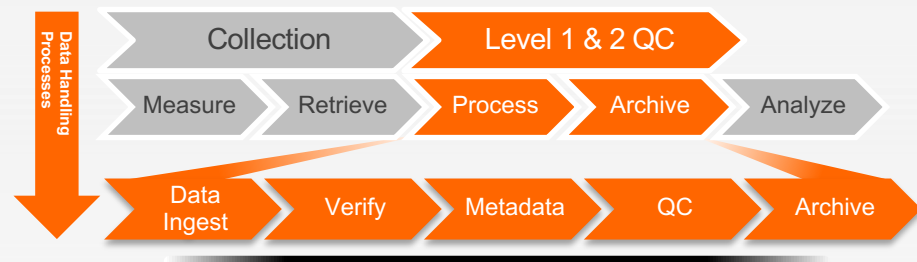
Eugene F. Burger<sup>1</sup>, Benjamin Pfeil<sup>3</sup>, Kevin O'Brien<sup>2</sup>, Linus Kamb<sup>2</sup>, Steve Jones<sup>3</sup>, Karl Smith<sup>2</sup>

<sup>1</sup>NOAA/PMEL, Seattle, WA; <sup>2</sup>University of Washington/JISAO, Seattle, WA; <sup>3</sup>Bjerknes Climate Data Centre (BCDC), Bergen, Norway

## The Data Processing Gap

Data assembly in support of global data products, such as GLODAP, and submission of data to national data centers for long-term preservation, demand significant effort. Delays in data assembly can negatively affect the timely production of scientific indicators that are dependent upon these datasets and data products.

What if data submission, metadata assembly and quality control can be combined into a single application? To support more streamlined data management processes NOAA's, Pacific Environmental Laboratory (PMEL), with support for the NOAA Ocean Acidification Program (OAP), and the Bjerknes Climate Data Centre (BCDC) within the Bjerknes Centre for Climate Research (BCCR) developers are developing such an application. This application has the potential for application towards a broader community, including the GLODAP collaborators.



## 1. Easy Data Ingest and Data Check

### 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**.

The data ingest dashboard

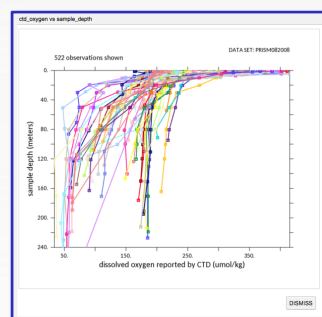
### Built-in Data Sanity Check

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. This development is lead by the Bjerknes Climate Data Centre (BCDC) developers.

Errors detected by the Sanity Check are highlighted

### Pre-QC Data Preview

A collection of preview plots 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 that the user can correct and resubmit the data. This step **improves data quality** by reducing common data mistakes before the data can be submitted to more rigorous quality control.



An example of a preview plot, before data are submitted for quality control

## Goal

These tools **streamline OA data processing, quality control and archival** by bridging the data workflow gap between data collection and data archival of biogeochemical data and metadata used by researchers. **These tools add value to the data by delivering high quality datasets.**

This application extends the web-based tools developed for SOCAT with a richer feature set applied to a broader range of biogeochemical variables that are measured by the Ocean Acidification research community. The workflow will contribute to the timely production of scientific indicators that are dependent upon these datasets, including synthesis products such as GLODAP.

## 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 use of these high-value data. These data processes will help scientists meet their obligations for data documentation, data access, and archival.

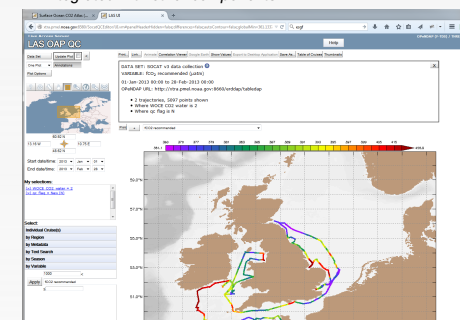
## 2. Integrated Metadata Entry Tool

Where possible, metadata are extracted from uploaded data. These extracted metadata are pre-populated in the metadata tool integrated with 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

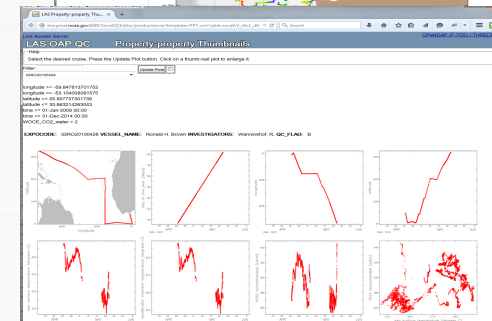
## 3. Quality Control Console

Quality control functionality being added to the dashboard allows the user to interactively review and set data quality flags for selected data points for a subset of biogeochemical variables.



## 4. 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.



At right: The quality control functionality that will be extended to incorporate a broader range of data used with ocean acidification research

## Contact

Eugene F. Burger, eugene.burger@noaa.gov, +1 206.526.4586

Benjamin Pfeil, Benjamin.Pfeil@uib.no, +47 55 58 98 39