

Streamlining Oceanic Biogeochemical Dataset Assembly in Support of Global Data Products

Eugene F. Burger¹, Benjamin Pfeil³, Kevin O'Brien², Linus Kamb², Steve Jones³, Karl Smith² ¹NOAA/PMEL, Seattle, WA; ²University of Washington/JISAO, Seattle, WA; ³Bjerknes 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) od Excel format. The data submission tool recognizes frequently submitted variables and identifies these in submitted data. This **allows ease and flexibility in data submission**.

Built-in Data Sanity Check

With the data properly identified, the sanity check warns the user if data are outside the bounds of preset 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.

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.

Contact

Eugene F. Burger, eugene.burger@noaa.gov, +1 206.526.4586 Benjamin Pfeil, Benjamin.Pfeil@uib.no, +47 55 58 98 39

Ann @ 15	e Grain Ira	e	ED Universit								
										Legge	d in an Ukar
•										Preferences Deni Feedback	Legest
Upter Desets	(see		Onteret ID	Observation Type	* Upload Date	Onto Statue	Meadata	Supplemental Documents	Anthinal	Fiename	
locate Columna			B04308	MONE	2010-11-08	RC-DARKer	2022/2021/05/	its.douroetbi	Sit specified	612,0086002.mm	
Columna Receipt	2		304000140606	MONUE	2010-11-05	Less	2019-12-03-20.00	(br. deserverb)	No.semified	V0018,66,030307.m	
THE OCH	- 2		DATIVENOCITIESU	CT HUR	2010-00-15	Cannol sheek	205-5-0201	Dr. Boursettil	Sanite 215-15 2018 N.M.M.	Oala, hemoiyee, aka	
openentar locuments			40092010810	PROFILE	2010-30-16 18-80-54 +0000	Leadings	2019/1/2022/01	IN AN AR	Samiller 2016.15 27 12-032 DMT	49,9210093-04.04	
Penten	. 6		ARGOMAR	orway	2019-08-13 23.15-39-4000	Cernel sheck	202.22.22.08.07	MARK AND A REAL PROPERTY.	Intentine 2019-11: 2019/2019/2019	872.40F	
NOMIN				CONTRACTOR	2019-08-10 2372:04 -0000	Carrol shark	2019-33-34 (808) /0022	in American	Not sensibled	1088,816Colt,32-88-287 map	
A214			MOGDIFFEDERED	CT HER	2010-01-02 10:05:05:40000	Carrol sheak	(or relaish)	(br. downwith)	Sit senifiel	INter 2011In2013.abs	
Fitter Detrainto			BATA/08/2011/2011	DOM:N	2010-01-02 10:34:04-0000	Carroi stask	2039-05-02.00.00	51. Sel Leetre, PAG, Fam. Inc. on Delet, SCLV, Dochese, 22003, Munif John Fax, Nat. 2007	Sandhe 2019-05 SC 30-4632 GAT	Onto, Test, 2011add Glasher	
Delawis			BATIMUSERHITCER	orway	2010-01-02 18/20-01-40000	Cernst sheck	its.missiski	its.douroetbi	Statested	Outs_Wile_2011to2013.stor	
Detek Detainto	10		DERCHIEARING	COMER .	2019-07-01 22:55-25-40000	Carrol shade	22/19/19/22 19:20	chargestat and catholic and catho	Same 215-05 12 YO M ST GAT	Outstate State Street and	LOBO I
			PREMORIZON	TRAECTORY	2010-11-21	Boneticer.	21546312159	DL BALENDS	Samme 2016-16 History Cont	PROPOSITION	
			PRIMOZODOBADOKI	PROFILE	2018-11-10 10-20-34 +0000	74100	2019/11/2012/07	senical Ami Antana sacime	Submitter 2018-15 19 1028-27 EMT	PROFESSIONALIN	
	- 15		NEHOPOETOENIN	monut	2018-11-08	Leacher	2010-12-08 (80.07	its.dourtectsi	Admint 2139-32	NEMOPOSTREEP 2.8V	

The data ingest dashboard



Errors detected by the Sanity Check are highlighted



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

3. Quality Control Console

Where possible, metadata are extracted from uploaded data. These extracted metadata are prepopulated 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.

Quality control functionality being added to the

dashboard allows the user to interactively review

and set data quality flags for selected data points

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

At right: The quality control functionality that will be extended to

research

incorporate a broader range of data used with ocean acidification

for a subset of biogeochemical variables.

4. Streamlined Data Archival

destination.







BCDC Bjerknes Climate Data Centre