

Endurance Array Current Data and QA/QC Activities

Christopher Wingard, Craig Risien, Russell Desiderio, Stuart Pearce



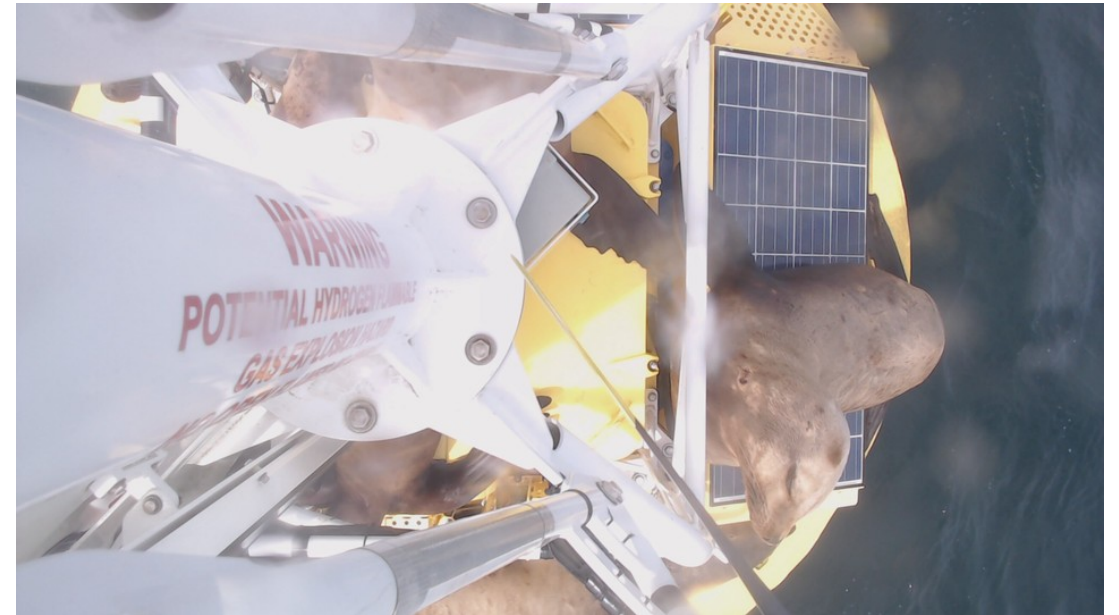
Endurance Data Team

Data Team Lead

- **Christopher Wingard** – Data transfers, ingestion, QA/QC, annotations, and asset management.

Data Team

- **Craig Risien** – QA/QC, in-depth assessments, and communications.
- **Russell Desiderio** – asset management, QA/QC, and in-depth assessments.
- **Stuart Pearce** – Glider operations and data management.



Spending ~1.0 FTE effort on Data QA/QC in
OOI 2.0 (Oct 2018-Apr 2019)

Endurance Data Team Responsibilities

- Asset Management
 - Entering required metadata (cal, deployment info) into OOINet.
- Data Ingestion/Availability
 - Ingesting/re-ingesting data into OOINet.
 - Ensuring data flow from platform to shore to science users.
- Data QC
 - Identifying/resolving issues with instruments/data.
 - Implementing automated and HITL data QC.
- Annotation
 - Annotating data in OOINet for instrument/data issues, metadata changes.
- Supporting CI (working groups, enhancements) and users (Help Desk)

Asset Management

- Documenting and formalizing asset management processes
 - EA began creating scripts in Fall 2015 for accuracy and efficiency.
 - Scripts have evolved as program needs have changed.
 - Instituted formal, two-person check of all metadata entered for Endurance 11 (Spring 2019).
 - Collaborating on documentation of SOPs and code used, including sharing and documenting automated processes for capturing calibration coefficients.
- Critical Metadata review of all past Endurance deployments
 - Created formal process and code for review.
 - Endurance 10 (Fall 2018) review completed; working backwards.
 - Will complete by end of July 2019.

Data Ingestion/Availability (Time Availability)

- Availability Assessment for 1.0 and 1st Quarter of 2.0
 - Developed code to assess data availability in OOI via M2M interface, cross-comparing with data available in internal OMS++ system.
 - Identified and backfilled several datasets, complex iterative process involving:
 - Identifying issue/causes for missing data
 - Determining corrective action
 - Implementation of corrective action
 - Re-running analysis
 - Example: METBK – Identified 125 days of missing data (2.46% of total) due to parser bug. Purged and re-ingested data with updated parser.
 - Example: NUTNR – Identified 1365 days of missing data (16.72% of total) due to missing parser (new instrument version). Setup ingests and ran with new parser.
 - Work will be completed by end of PY1.
 - Code shared with all MIOs.

Data Ingestion/Availability (CE02SHSM)

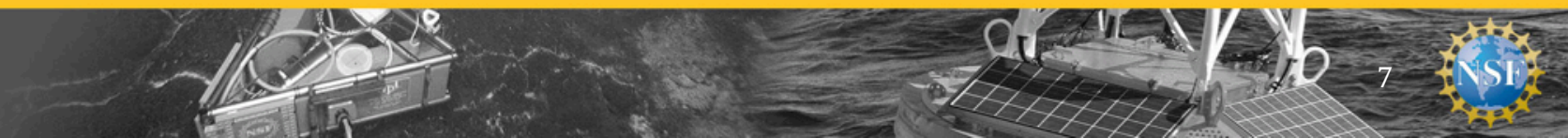
CE02SHSM	UFrame	OMS++	% Difference	Start Date	End Date	Total Days	Difference (days)	Ingest ID	Comments
FDCHP BUOY	0.00	0.00	0.00	4/2/2015	9/30/2018	1277	0.00	N/A	There was an annotation associated with the FDCHP that prevented any data from being downloaded from uFrame. I cleared that annotation and set the appropriate QA/QC flag on the motion-corrected and flux measurements. The non-corrected data is OK and should be available to interested users.
METBK BUOY	95.23	96.09	0.86	4/2/2015	9/30/2018	1277	10.99		
PCO2A BUOY	96.24	96.24	0.00	4/2/2015	9/30/2018	1277	0.00		
VELPT BUOY	96.32	96.32	0.00	4/2/2015	9/30/2018	1277	0.00		
WAVSS BUOY	90.77	90.77	0.00	4/2/2015	9/30/2018	1277	0.00		
ADCPT NSIF	94.60	0.70	-93.90	4/2/2015	9/30/2018	1277	-1199.06		
CTDBP NSIF	86.15	94.91	8.76	4/2/2015	9/30/2018	1277	111.91	13840	There was an issue with power-cycling DCL27 during D00007. This resulted in gap in data collection that exceeded the telemetered ingest timeout window (30 days). Once we fixed the issue and data started flowing again, the ingests should have been restarted. I re-ran the ingests for this deployment and have filled in the gap.
DOSTA NSIF	86.15	94.91	8.76	4/2/2015	9/30/2018	1277	111.91	13841	There was an issue with power-cycling DCL27 during D00007. This resulted in gap in data collection that exceeded the telemetered ingest timeout window (30 days). Once we fixed the issue and data started flowing again, the ingests should have been restarted. I re-ran the ingests for this deployment and have filled in the gap.
FLORT NSIF	73.63	83.41	9.78	4/2/2015	9/30/2018	1277	124.90	13843	There was an issue with power-cycling DCL27 during D00007. This resulted in gap in data collection that exceeded the telemetered ingest timeout window (30 days). Once we fixed the issue and data started flowing again, the ingests should have been restarted. I re-ran the ingests for this deployment and have filled in the gap.
NUTNR NSIF	71.91	94.21	22.30	4/2/2015	9/30/2018	1277	284.78	13844	There was an issue with power-cycling DCL27 during D00007. This resulted in gap in data collection that exceeded the telemetered ingest timeout window (30 days). Once we fixed the issue and data started flowing again, the ingests should have been restarted. I re-ran the ingests for this deployment and have filled in the gap. Additional gap due to transition to SUNA starting with D0007
OPTAA NSIF	79.81	49.61	-30.20	4/2/2015	9/30/2018	1277	-385.68	13842	There was an issue with power-cycling DCL27 during D00007. This resulted in gap in data collection that exceeded the telemetered ingest timeout window (30 days). Once we fixed the issue and data started flowing again, the ingests should have been restarted. I re-ran the ingests for this deployment and have filled in the gap.
PHSEN NSIF	83.26	82.24	-1.02	4/2/2015	9/30/2018	1277	-12.99		
SPKIR NSIF	98.44	98.51	0.08	4/2/2015	9/30/2018	1277	1.00		
VELPT NSIF	98.36	98.36	0.00	4/2/2015	9/30/2018	1277	0.00		

Data Ingestion/Availability (Glider DAC)

“The mission of the Glider DAC is to provide glider operators with a simple process for submitting glider data sets to a centralized location, enabling the data to be visualized, analyzed, widely distributed via existing web services and the Global Telecommunications System (GTS) and archived at the National Centers for Environmental Information (NCEI).”

<https://gliders.ioos.us/data/>

- Assumed responsibility (Stuart Pearce) in 2.0 for transmitting OOI Endurance glider data to the Glider DAC.
- Received Glider DAC code which needed new code to add additional science data variables (new feature coming to the Glider DAC).
- Glider DAC code for Endurance OOI glider data is complete and running. All Endurance glider deployments will be uploaded to the Glider DAC by end of month, May 2019.
- Automating real time processing and uploading of data to the Glider DAC in progress.
- Working to generalize code and documentation for Pioneer, Global and possibly other glider users.

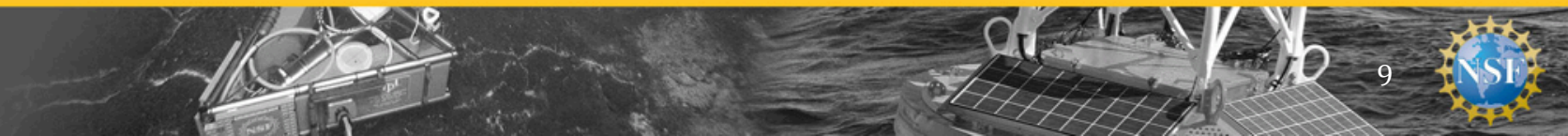


Data QA/QC

- Operations and Data Monitoring: Implementing formal bi-weekly field operations review, assignment of engineering & data actions, and updating OMS++ infrastructure and instrument alarming. Informal daily assessments.
- Addressing/identifying issues via OOI Data Team working groups (QARTOD, ADCP, Water Sampling, Ticketing, etc.).
- Developed tools for pulling data from M2M to create different visualizations for HITL QA/QC (built off Time Availability Assessments).
- Mooring meteorological and surface wave data pushed to NDBC. Automated routines compare data to nearby source(s):
 - Station IDs: 46097 (CE02SHSM), 46098 (CE04OSSM), 46099 (CE07HSM) and 46100 (CE09OSSM).

Data QA/QC

- Developed code to process data MMP datasets, creating single merged datasets. Will be used to generate NetCDF files for OOI ERDDAP server.
- Updating code used to generate datasets for OMS++ to include ADCP data from all moorings, and adding all inductive modem hosted instruments.
- Completed in-depth analysis of uncabled profiler systems, identifying critical errors in treatment of source data and (time and pressure) and merging of datasets. Redmine tickets generated.



Endurance Accomplishments To Date

- Transitioning roles and responsibilities within group to adapt to new role.
- Attended OOI Data Team Workshop at UW in January 2019.
- Completed re-ingestion of backlog assigned by 1.0 Data Team and have made significant progress toward completion of data availability assessment.
- Participated in creation of a common format/spreadsheet for OOI Water Sampling data across the program.
 - All past data is processed and entered into standardized spreadsheets with exception of salinity samples.
 - Will complete final, “second set of eyes” review of spreadsheets by end of June 2019 with upload to Alfresco as complete.
- Successful Asset Management updates and Data Ingestion for:
 - Endurance 10 (Fall 2018); Endurance 11a, 11b (Spring 2019); multiple glider and CSPP cruises

Endurance Plan Going Forward

- Bi-Weekly meetings to identify tasking for Mooring and Data Operations.
- Complete critical metadata review, end of July 2019.
- In-depth review of existing annotations.
- Evolution of OMS++ system to improve mooring and data monitoring (alerts & alarms, data visualization).
- Engaging SMEs to confirm/determine best practices and data review methodology.
- Continue to develop and implement automated processes for data reviews.
- Continue OOI Data Team work on program-wide issues/processes.