

# OOI Data Quality Procedures and Tools

M. Vardaro, M. Crowley, L. Belabbassi, L. Garzio,  
J. Kerfoot, F. Knuth, S. Lichtenwalner, M. Smith

**Rutgers University, New Brunswick, NJ**



Ocean Hackweek 2018

1



# Overview

1. Data Flow & Products
2. Data Review Procedures
3. Periodic Reviews & Documentation
4. Next Steps
5. Conclusions



# OOI By the Numbers



**7** Arrays

**57** Stable Platforms  
Moorings, Profilers, Nodes

**31** Mobile Assets  
Gliders, AUVs

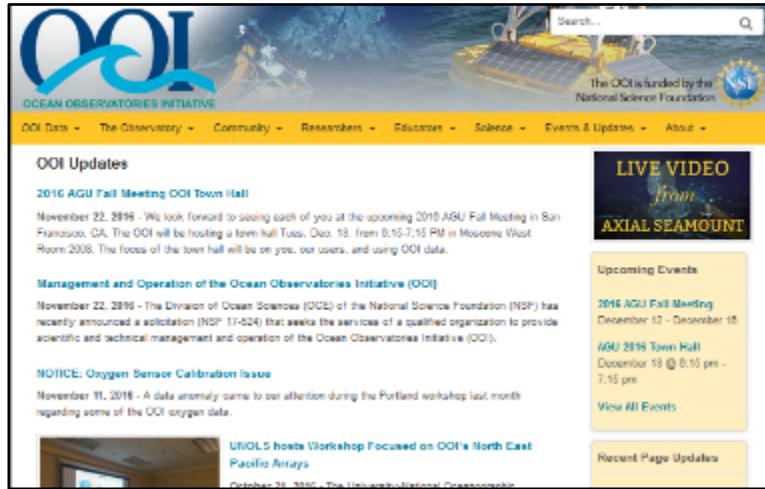
**1227** Instruments (~850 deployed)

**>2500** Science Data Products

**>100K** Science/Engineering Data Products



# OOI: Web Portals



Main

**Research Arrays** Select an array on the map or choose from the list.

- Endurance
- Station Papa
- Pioneer
- Argentine Basin
- Immergée Sea
- Southern Ocean
- Cabled Array

Home      Q Data Catalog Search      Glossary      FAQ      Log In

The map displays the world's oceans with green dots indicating the locations of various research arrays. Labeled regions include the Arctic Ocean, North Pacific Ocean, South Pacific Ocean, North Atlantic Ocean, South Atlantic Ocean, and Southern Ocean. The map also shows coastlines and major rivers.

ooNet

The screenshot shows the homepage of the Ocean Education Portal. At the top, there's a navigation bar with links for "About the OEP", "Learn to Use Our Tools", "Teach with Tools", and "My Profile". Below the header, there's a main content area with a "Welcome" section featuring a video thumbnail of two people at a computer. There are also three cards below: "Mission Data" (with a line graph thumbnail), "Connected Concepts" (with a group of people thumbnail), and "Current Investigations" (with a person in a lab thumbnail). The overall design is clean with a blue and white color scheme.

Shipboard Data

Company Home > OOI > Cabled Array > Cruise Data

**Cruise Data**  This view allows you to browse the items in this space. Shipboard Data from Cabled Array Cruises

(0)    More Actions 

▼ Browse Spaces Items Per Page: 200

Name	Description	Created	Modified	Actions
Cabled_1_TN_221_2008-7-22		28 October 2015 14:15	28 October 2015 14:15	 
Cabled_2_TN_252_2010-7-26		28 October 2015 14:15	28 October 2015 14:15	 
Cabled_3_TN_268_2011-8-11		28 October 2015 14:27	28 October 2015 14:27	 
Cabled_4_TN_299_2012-06-30	2013 OOI Cabled Array deployment cruise, R/V Thompson (R299), June 30-August 23, 2013	24 October 2015 16:00	16 May 2016 15:40	 
Cabled_5_TN_313_2014-7-13		28 October 2015 15:00	28 October 2015 15:00	 
Cabled_6_TN_326_2015-7-04		28 October 2015 15:16	28 October 2015 15:16	 

Page: 1 of 1    

▼ Content Items Items Per Page: 200

Name	Description	Size	Created	Modified	Actions
OOS_Store		4 KB	16 May 2016 10:13	16 May 2016 10:13	 
OOS_Store		6 KB	16 May 2016 10:13	16 May 2016 10:14	 

Page: 1 of 1    

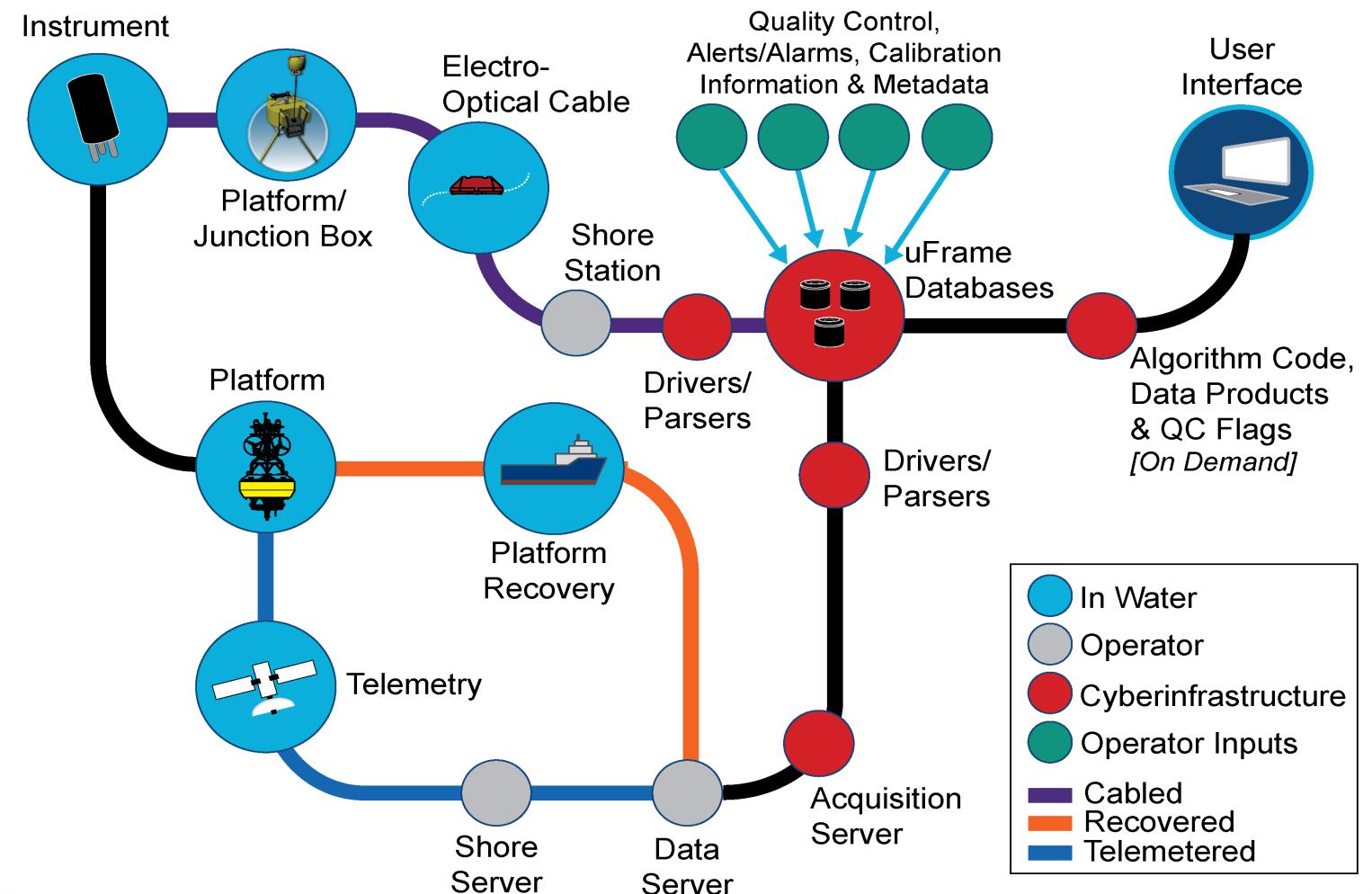
# Data Flow Chart

## Data types:

- Telemetered Data
- Recovered Data
- Streamed Data
- Shipboard Data
- Metadata (CF 1.6)

## Data product levels:

- Raw, L0, L1, L2



# OOI Data Product Levels

- **Raw data:** The datasets as they are received from the instrument
  - May contain multiple L0, L1, or L2 parameters, data for multiple sensors, and be in native sensor units
  - Always persisted and archived by the OOI
  - Example: format 0 binary file from an SBE-37IM on a Global Flanking Mooring.
- **Level 0 (L0):** Unprocessed, parsed data parameter that is in instrument/sensor units and resolution
  - Sensor by sensor (unpacked and/or de-interleaved) and available in OOI supported formats (e.g., NetCDF)
  - Always persisted and archived by the OOI
  - Example: SBE-37IM Temperature portion of the hex string
- **Level 1 (L1):** Data parameter that has been calibrated and is in scientific units
  - Created on-demand via user synchronous or asynchronous request
  - Actions to transform Level 0 to Level 1 data are captured and presented in the metadata of the Level 1 data
  - Example: SBE-37IM Temperature converted from hex to binary and scaled to produce degrees C
- **Level 2 (L2):** Derived data parameter created via an algorithm that draws on multiple L1 data products
  - Created on-demand via user synchronous or asynchronous request
  - Products may come from the same or separate instruments; data from all instruments are provided during download
  - Example: SBE-37IM Density and Salinity



# OOI Data Status

- All available telemetered and recovered data are being ingested and made accessible for download
- Rutgers CI currently storing >340TB of raw data, including ~100 TB HD video (so video ≈ 30% of total data)
  - ~150TB are publicly available via the raw data archive
  - ~14TB of processed data ready for product generation
- All known issues have been annotated. Unless otherwise remarked, the data have not been validated against *in situ* samples taken by instrumentation external to the system.

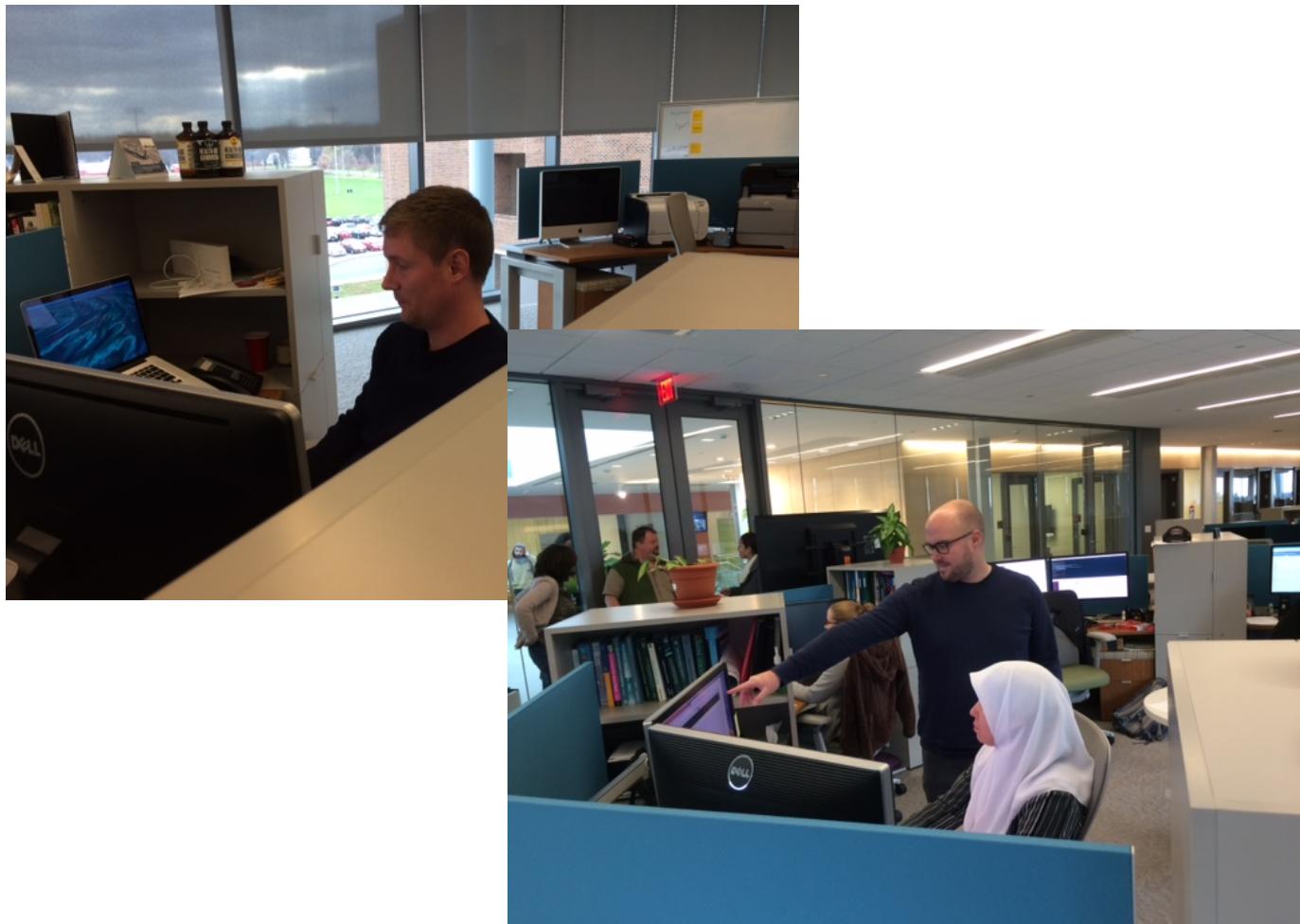


# Overview

1. Data Flow & Products
2. Data Review Procedures
3. Periodic Reviews & Documentation
4. Next Steps
5. Conclusions



# Data Team Primary Goals



1. Monitor the operational status of data flowing through the OOI system
2. Ensure the availability of OOI datasets in the system (raw, processed, derived, and cruise)
3. Ensure that data delivered by the system meets quality guidelines
4. Report operational statistics on data availability, data quality, and issue resolution



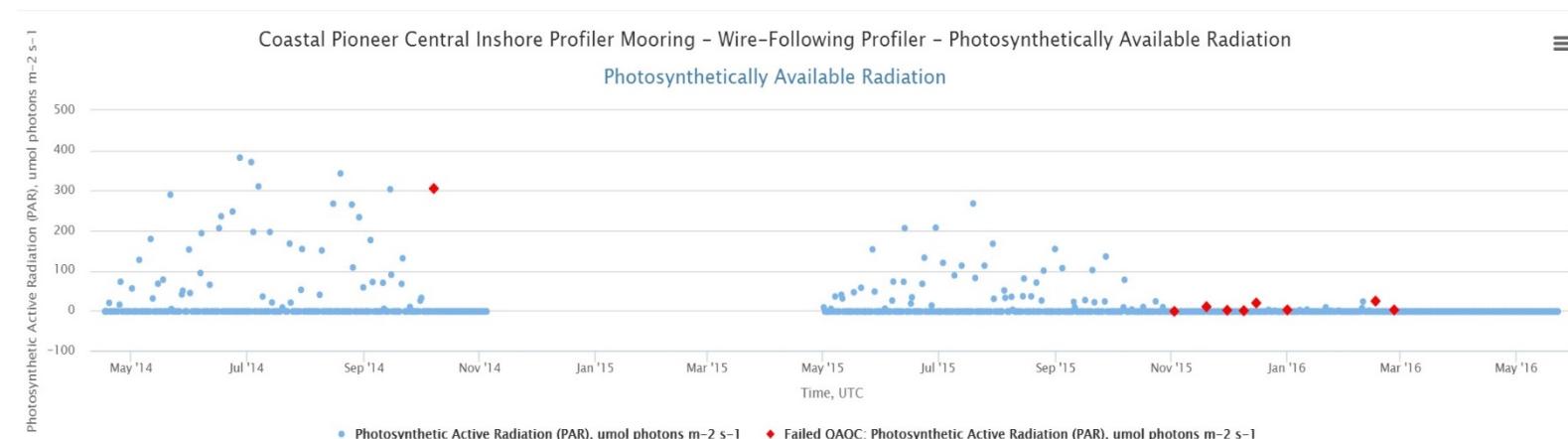
# Data Team Tasks



# OOI Automated QC Procedures

- 6 automated QC algorithms can produce 7 flags (including logical “or” which combines flags) which are plottable and are included in downloaded files
- Coded based on specifications written by OOI Project Scientists, derived from QARTOD manuals and other observatory experiences
- Algorithms refer to “lookup tables” assembled by OOI Project Scientists with input from subject matter experts: <https://github.com/ooi-integration/qc-lookup>

1. Global Range Test
2. Local Range Test
3. Spike Test
4. Stuck Value Test
5. Trend Test
6. Temporal Gradient Test
7. Spatial Gradient Test (Profile)



# NetCDF QC flags

**NOTE:** Results of the QC tests do not remove data.

- Each parameter in a file has a corresponding \*\_qc\_executed and \*\_qc\_results variable
  - qc\_executed indicates which tests were run
  - qc\_results denotes which tests passed
- Integer values can be converted to a binary that indicates which tests were run (1) or not (0), and which passed (1) or failed (0).
- See example QC Python notebook or online guide (<http://oceanobservatories.org/knowledgebase/interpreting-qc-variables-and-results/>) for more details.



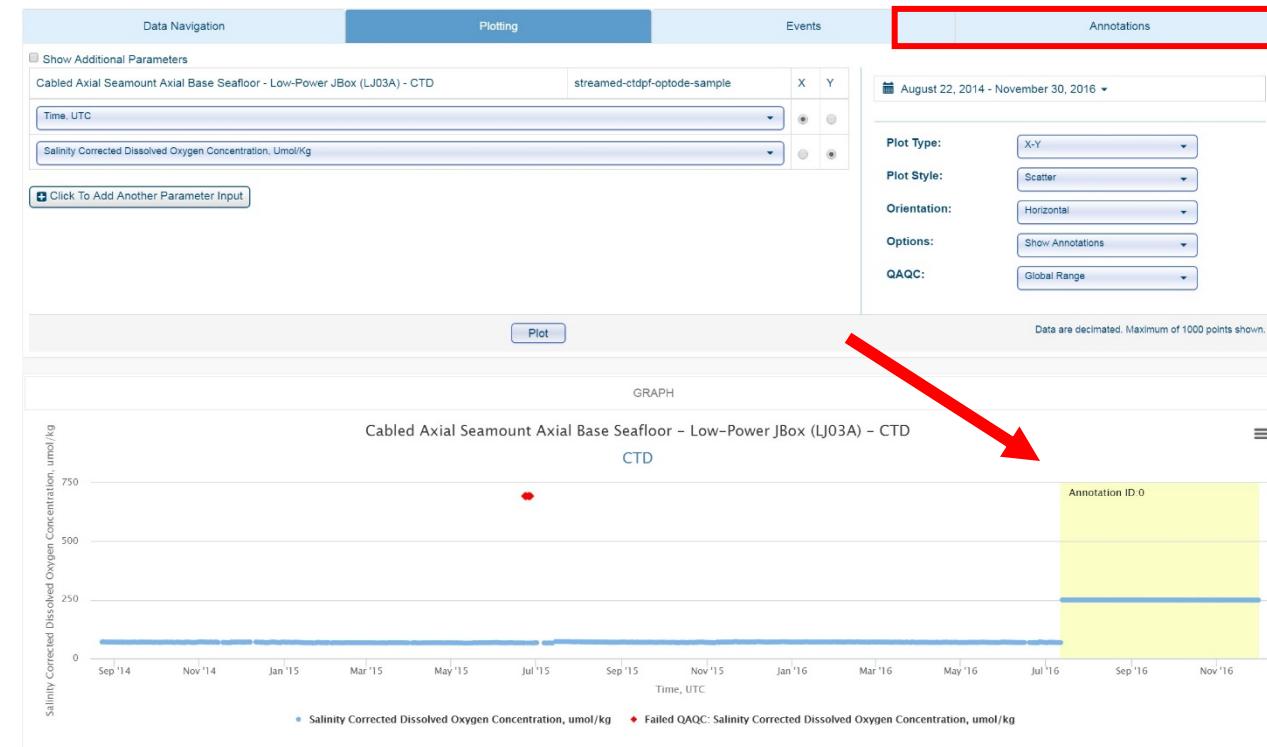
# Data Status Categories

Status	Description	QARTOD Code	QARTOD Description	Color
NOT_OPERATIONAL	Instrument not functional (no data expected)		Not operational	#667380
NOT_AVAILABLE	Instrument functional, data lost in transmission	9	Missing data	#A9A9A9
PENDING_INGEST	Instrument functional, data exists, Awaiting ingest			#F0F0F0
NOT_EVALUATED	Instrument functional, data exists, Awaiting evaluation	2	Not evaluated, not available or unknown	#003366
SUSPECT	Instrument functional, data exists and either failed a QC test or does not reflect environmental conditions	3	Questionable/suspect	#C8A23D
FAIL	Instrument functional, data exists but is known to be bad due to known instrument or calibration error	4	Bad	#FF0000
PASS	Instrument functional, data exists, passed QC tests, is complete and looks reasonable	1	Good	#008000
GOOD	Instrument functional, data exists, passed QC tests, is complete and has undergone validation with shipboard datasets and reached the highest level of QC that the OOI can provide			#006400



# Data Annotation

- Annotations are the primary means of communication between data team and users
- Annotations can be directly entered via the GUI for specified data streams
- Annotation text appears in a tab on the data catalog/plotting page
- Annotation time ranges can be shown on plots (via “Options” interface)
- Annotations also included in downloaded data



Annotation ID	Annotation	Reference Designator	Stream Name	Start Date	End Date	Exclude Data?
0	These data are suspect, possibly due to incorrect vendor calibration values. Raw phase data should be correct, but the derived O2 products should not be used from 7/12/16 onwards.	RS03AXBS-LJ03A-12-CTDPFB301	streamed_ctdpf-optode-sample	Tue, 12 Jul 2016 00:00:00 GMT	Thu, 01 Dec 2016 23:41:00 GMT	false



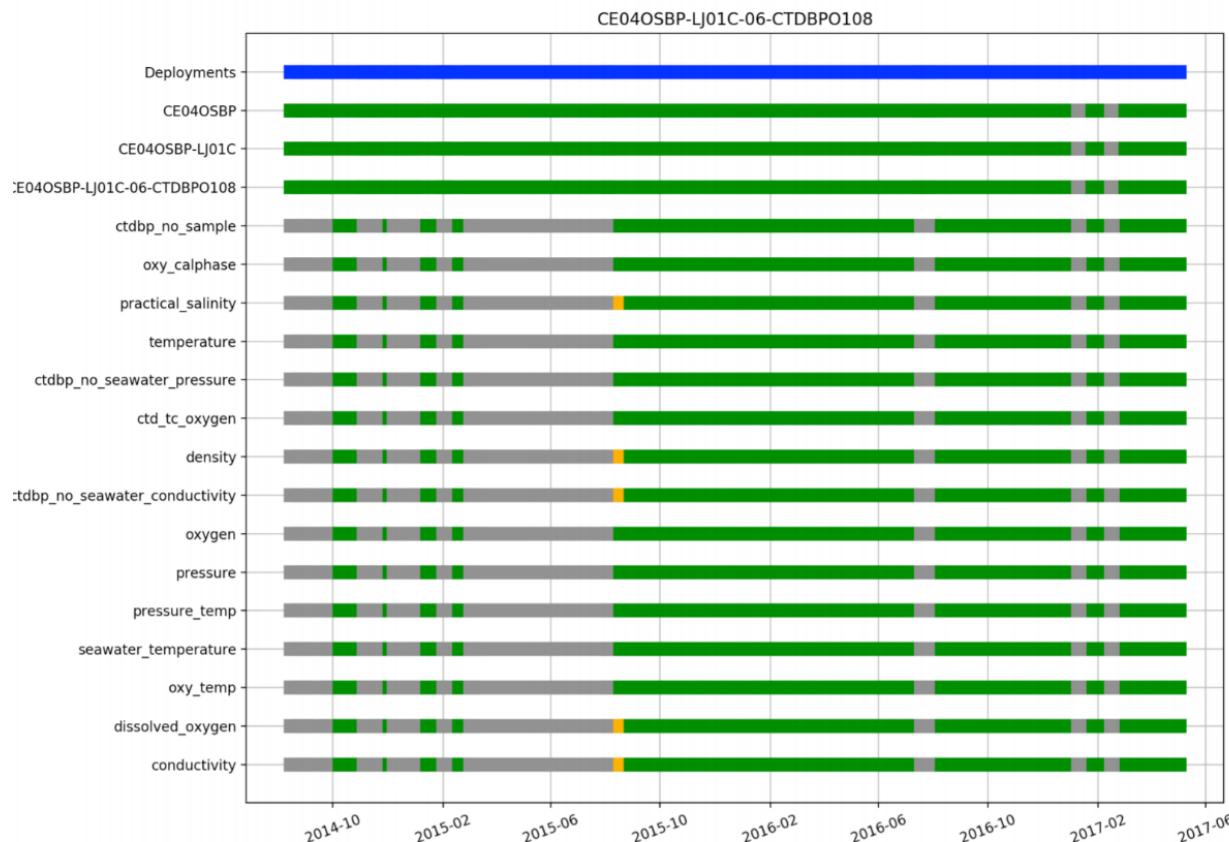
# Overview

1. Data Flow & Products
2. Data Review Procedures
3. Periodic Reviews & Documentation
4. Next Steps
5. Conclusions



# Reviews and Reporting

## Quality Timeline



## Annotation Text

Level	Deployment	StartTime	EndTime	Annotation	Status	Redmine#
ctdbp_no_sample	D00001	2014-08-15T00:12:00Z	2014-08-25T18:50:41Z		NOT_AVAILABLE	
ctdbp_no_sample	D00001	2014-08-31T19:13:27Z	2014-09-22T22:42:44Z		NOT_AVAILABLE	
ctdbp_no_sample	D00001	2014-11-04T16:05:51Z	2014-11-05T18:56:20Z		NOT_AVAILABLE	
ctdbp_no_sample	D00001	2014-11-14T18:36:38Z	2014-11-17T18:03:22Z		NOT_AVAILABLE	
CE04OSBP		2014-12-07T19:45:00Z	2014-12-16T00:00:00Z	PFE down. HVPS1 MOV explosion, 800A breaker tripped, investigation and restoration		12264
ctdbp_no_sample	D00001	2014-12-07T20:59:40Z	2014-12-16T22:29:37Z		NOT_AVAILABLE	
ctdbp_no_sample	D00001	2014-12-16T23:44:08Z	2014-12-29T20:30:01Z		NOT_AVAILABLE	
CE04OSBP		2015-01-07T07:32:00Z	2015-01-07T08:06:00Z	PNWGP Portland <-> Seattle outage		12264
CE04OSBP		2015-01-31T00:00:00Z	2015-02-04T00:00:00Z	Intermittent partial data loss due to storage drive problems at OTB		12264
ctdbp_no_sample	D00001	2015-01-31T23:59:59Z	2015-02-03T09:56:15Z		NOT_AVAILABLE	
ctdbp_no_sample	D00001	2015-03-03T02:16:22Z	2015-03-06T19:57:13Z		NOT_AVAILABLE	
ctdbp_no_sample	D00001	2015-03-06T19:58:48Z	2015-08-02T00:00:00Z		NOT_AVAILABLE	
CE04OSBP		2015-03-21T14:10:00Z	2015-03-22T04:20:00Z	PNWGP outage due to City of Seattle fiber cable work		12264
CE04OSBP		2015-06-13T00:00:00Z	2015-06-15T16:30:00Z	Network issues due to fire that damaged fibers between Portland and Seattle		12264
practical_salinity	D00002	2015-08-12T01:00:00Z	2015-08-14T02:00:00Z	Unusual drop in conductivity values.	SUSPECT	
density	D00002	2015-08-12T01:00:00Z	2015-08-14T02:00:00Z	Unusual drop in conductivity values.	SUSPECT	
ctdbp_no_seawater_conductivity	D00002	2015-08-12T01:00:00Z	2015-08-14T02:00:00Z	Unusual drop in conductivity values.	SUSPECT	
dissolved_oxygen	D00002	2015-08-12T01:00:00Z	2015-08-14T02:00:00Z	Unusual drop in conductivity values.	SUSPECT	
conductivity	D00002	2015-08-12T01:00:00Z	2015-08-14T02:00:00Z	Unusual drop in conductivity values.	SUSPECT	
CE04OSBP		2015-08-29T00:00:00Z	2015-08-29T00:30:00Z	Outage during major utility power failure in Seattle		12264
CE04OSBP		2016-01-07T06:10:00Z	2016-01-07T06:52:00Z	Four 1-minute outages between Portland and Seattle due to maintenance		12264
CE04OSBP		2016-03-03T23:06:00Z	2016-03-11T09:30:00Z	Fiber break between Portland and Seattle		12264
CE04OSBP		2016-05-20T16:33:00Z	2016-05-20T18:04:00Z	Fiber break between Portland and Pacific City		12264
CE04OSBP		2016-07-12T02:53:00Z	2016-07-12T03:51:00Z	Unexplained loss of power at Pittock Building in Portland		12264
ctdbp_no_sample	D00002	2016-07-18T00:42:58Z	2016-07-19T21:06:56Z		NOT_AVAILABLE	
ctdbp_no_sample	D00003	2016-07-22T22:50:00Z	2016-07-25T19:51:39Z		NOT_AVAILABLE	
CE04OSBP		2016-12-17T18:00:00Z	2016-12-17T19:00:00Z	Corvallis data center lost power		12264
CE04OSBP		2016-12-22T01:50:00Z	2016-12-23T12:44:00Z	Fiber break in Portland due to train crash		12264
CE04OSBP		2017-01-08T19:58:00Z	2017-01-08T21:41:00Z	Network outage during major Seattle utility power failure		12264
ctdbp_no_sample	D00003	2017-01-09T18:30:53Z	2017-01-11T01:16:53Z		NOT_AVAILABLE	
CE04OSBP		2017-01-09T18:32:00Z	2017-01-09T21:30:00Z	Lightning strike in Pacific City led to data interruption through isolation of both cable lines from shore station equipment.	NOT_OPERATIONAL	11776
CE04OSBP		2017-02-07T13:00:00Z	2017-02-07T15:00:00Z	Outage during PNWGP 1-hour router-maintenance		12264
ctdbp_no_sample	D00003	2017-02-15T14:43:05Z	2017-02-16T22:27:12Z		NOT_AVAILABLE	
CE04OSBP		2017-02-15T14:43:06Z	2017-02-15T17:16:00Z	On Wednesday, February 15, power to the North and South cable	NOT_OPERATIONAL	11998



# QC Database Tool ([ooi.visualocean.net](http://ooi.visualocean.net))

- Used for reference & statistics
- Includes status information, as well as a cruise data checklist
- Includes testing/review capability
- Annotation options

OOI Data Team

Arrays / ... / CE02SHBP-LJ01D-06-CTDBPN106 / streamed / ctldb\_no\_sample

OOI Site Page Data Portal

**Data Stream Report**

Instrument Name	CTD	Uframe Route	Driver
Reference Designator	CE02SHBP-LJ01D-06-CTDBPN106	Parser	seabird.sbe16plus_v2.ctldb_no.driver
Method	streamed	Instrument Type	Science
Stream	ctldb_no_sample		

**Data Availability Plot**  
from 9/10/2014 3:43:00 PM to 5/11/2017 5:15:30 PM

**Annotations** **Parameters**

**Annotations**

Metadata	Start Date	End Date	Comment
CE02SHBP-LJ01D-06-CTDBPN106	9/10/14, 3:43 PM	9/25/14, 6:17 PM	Exact cause of data gap at beginning of deployment currently unknown. Most likely due to testing as the instrument was coming online.  <b>Todo:</b> check: difference between data begin and deployment begin date is: 15 days 02:34:58 By friedrich, on 3/29/17
Deployment: 1			
Method: streamed			
Stream: ctldb_no_sample			

CE02SHBP-LJ01D-06-CTDBPN106

9/25/14, 6:17 PM    11/4/14, 4:05 PM    **Todo:** check: evaluate parameters  
By friedrich, on 3/29/17



# Data Availability

- QC Database: [ooi.visualocean.net](http://ooi.visualocean.net)

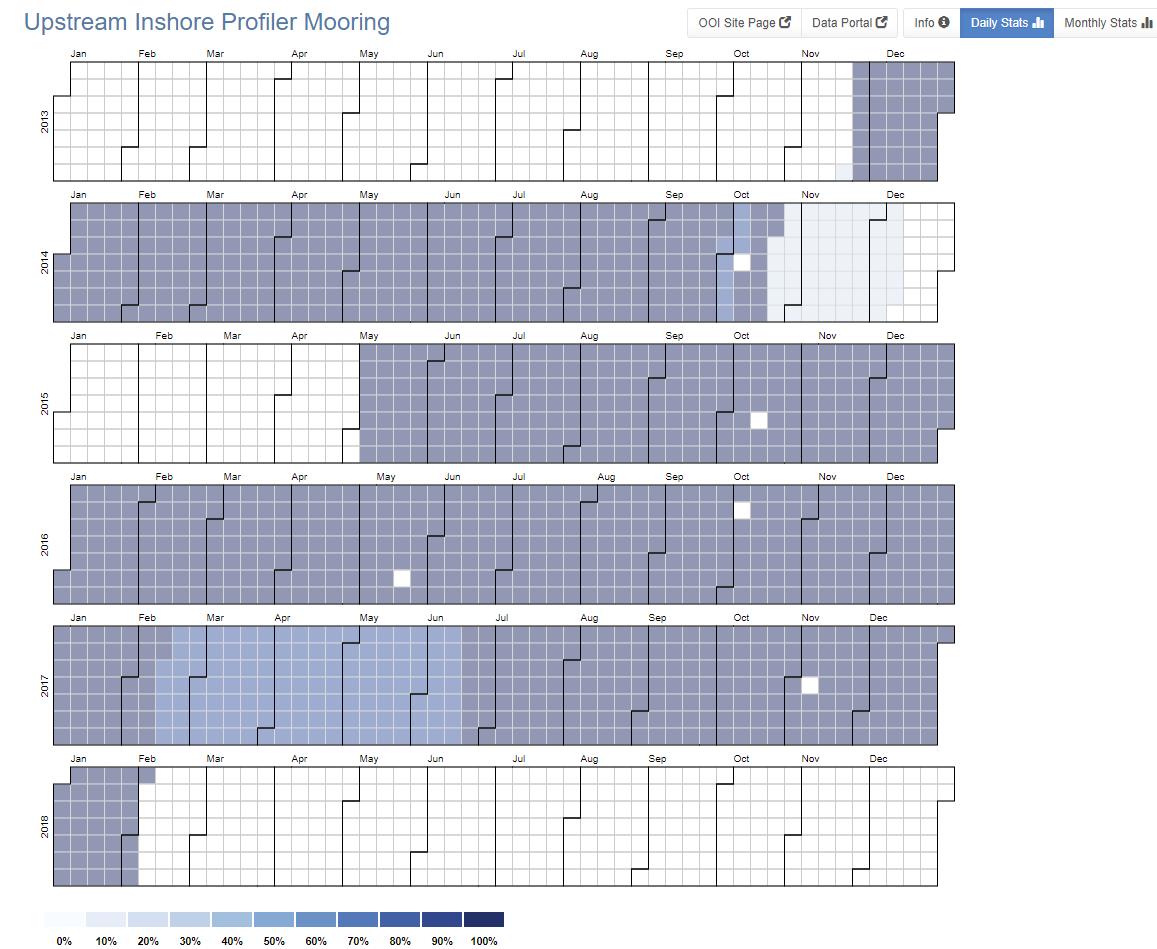
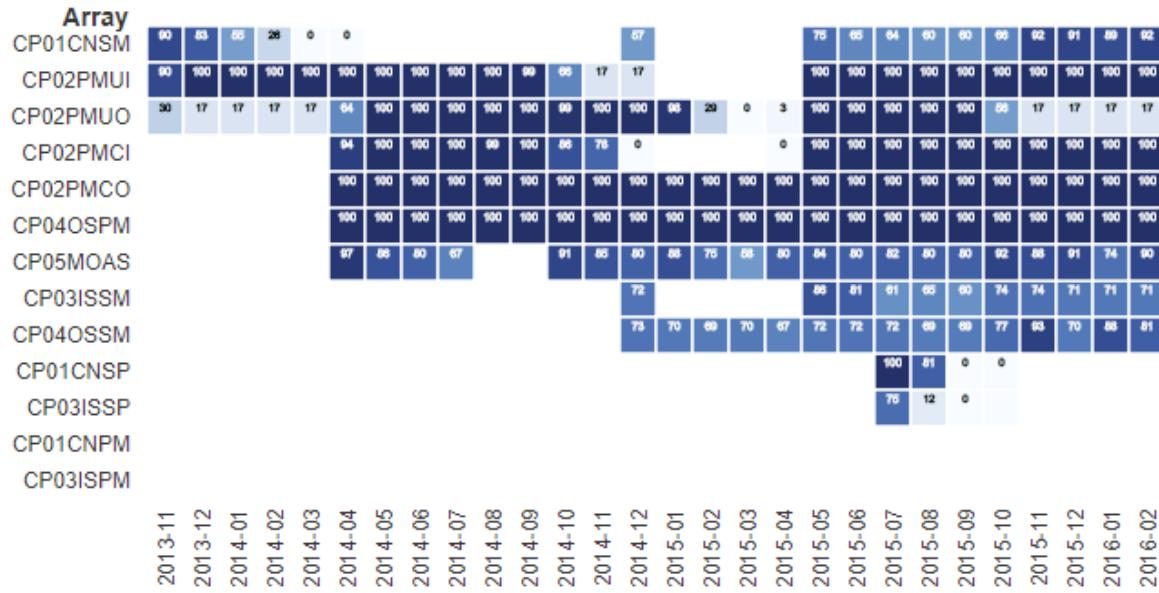
# Coastal Pioneer

Info  Daily Stats  Monthly Stats 

# Data Availability

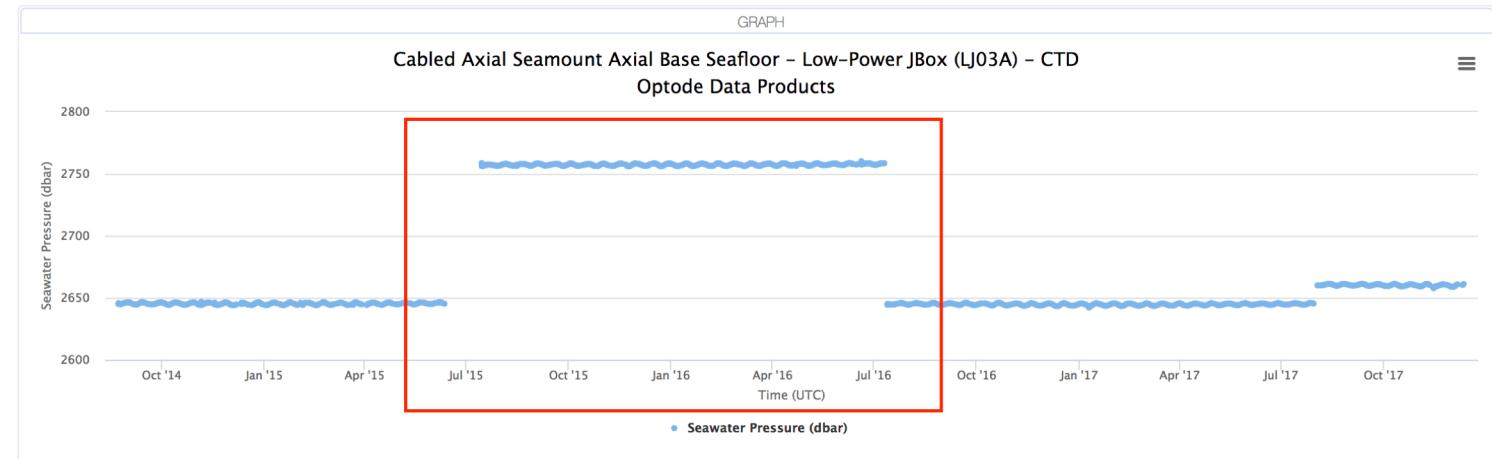
- QC Database: [ooi.visualocean.net](http://ooi.visualocean.net)

# Coastal Pioneer

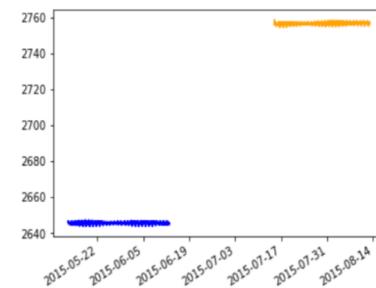
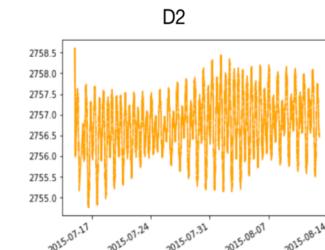
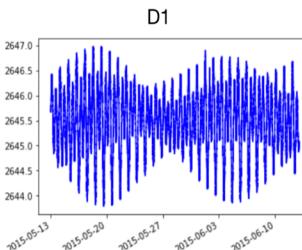


# Examples of QC Effort

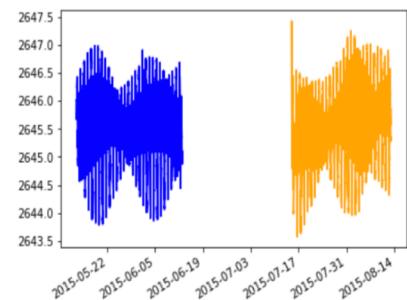
- Pressure offsets  
(Bad cal values)
- Drift
- Biofouling
- Damage
- Video issues



Ref Des: RS03AXBS-LJ03A-12-CTDPFB301  
Affected Deployment: 2  
UID: ATAPL-67627-00005  
SN#: 16-50128

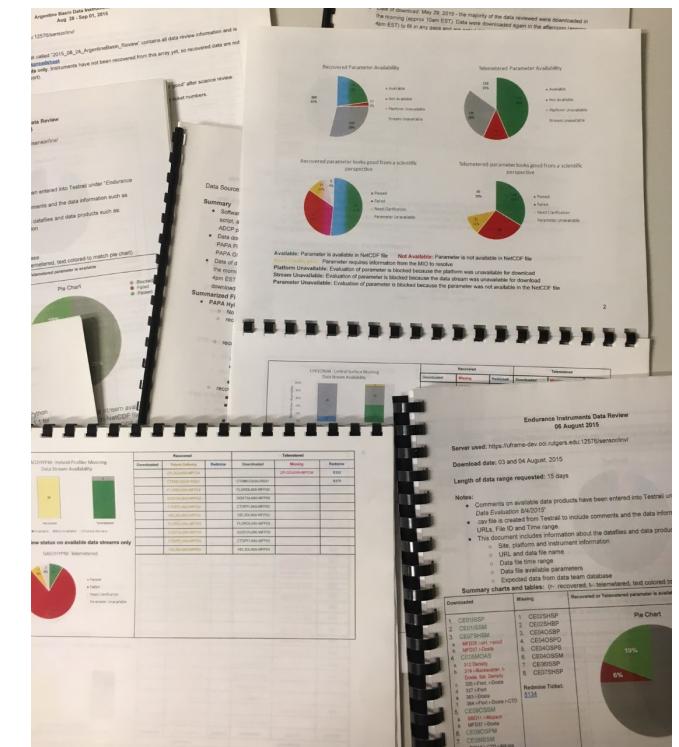


-111.19



# Deliverables

- Data Availability Reports
  - (% completeness, streams/parameters reported, particles in the system)
- Data Quality Reports
- Redmine reporting
  - Issues found, investigations, and Help Desk open/closed
- Deep dive investigation reports
- Annotations (to users)
- Download statistics



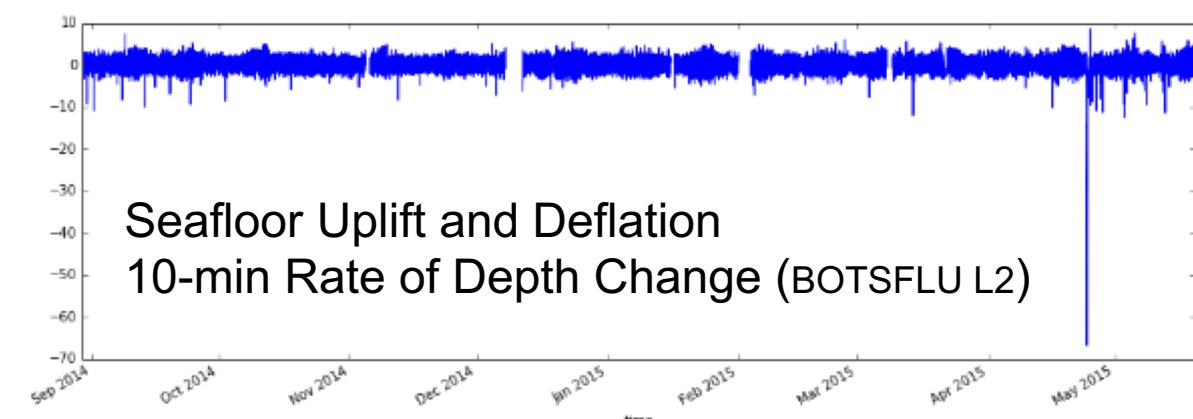
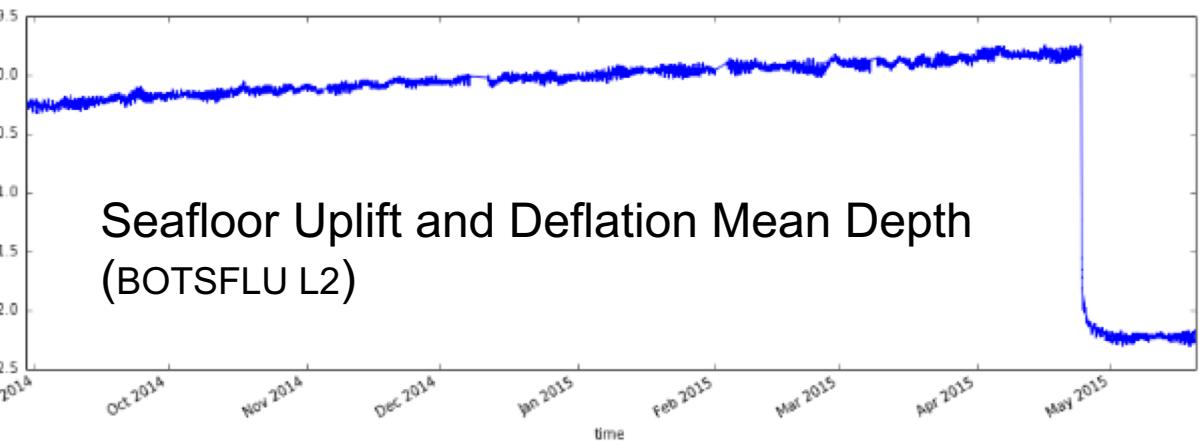
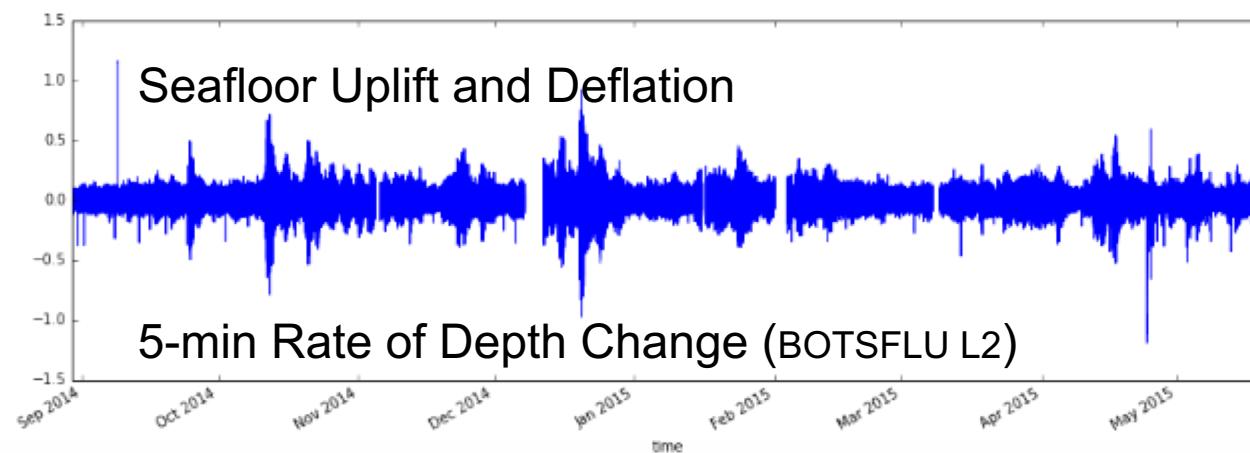
# Overview

1. Data Flow & Products
2. Data Review Procedures
3. Periodic Reviews & Documentation
4. Next Steps (into 2.0)
5. Conclusions



# Data Availability and Completeness

- Some derived data products are still being added to the system
- Some products are available but do not show up in data catalog or cannot be plotted
- Some products require additional processing (e.g. ZPLSC, HYDBB)



# Adding capability to OOI Net experience

Ocean Observatories » Web Interface

Overview Activity Roadmap Issues New Issue Gantt Calendar News Documents Wiki Files Settings

Enhancement #10913

Color bar needs to scale to majority of data for binned pseudocolor plots

Added by Daniel Maher 4 months ago. Updated 16 days ago.

Status: Ready for Work  
Priority: Urgent  
Assignee: Daniel Maher  
Category:  
Target version:  
Target Release:  
Category 1:  
Severity:  
Issue Closed:

Start date:  
Due date:  
% Done:  
Spent time:

20%

Array Affected:  
Instrument Affected:  
CI Software Affected:  
Work Breakdown Structure (WBS):

Northward UI Plot  
Velocity Profiler (150kHz)

Eastward UI Plot  
Velocity Profiler (150kHz)

Python Plots  
RS01SBPS-PC01A Stream: adcp\_velocity\_beam 2016.09.30T16.00.00 - 2016.10.07T16.00.00

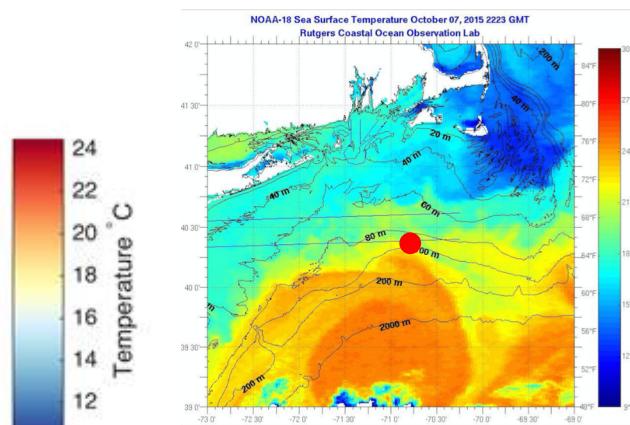
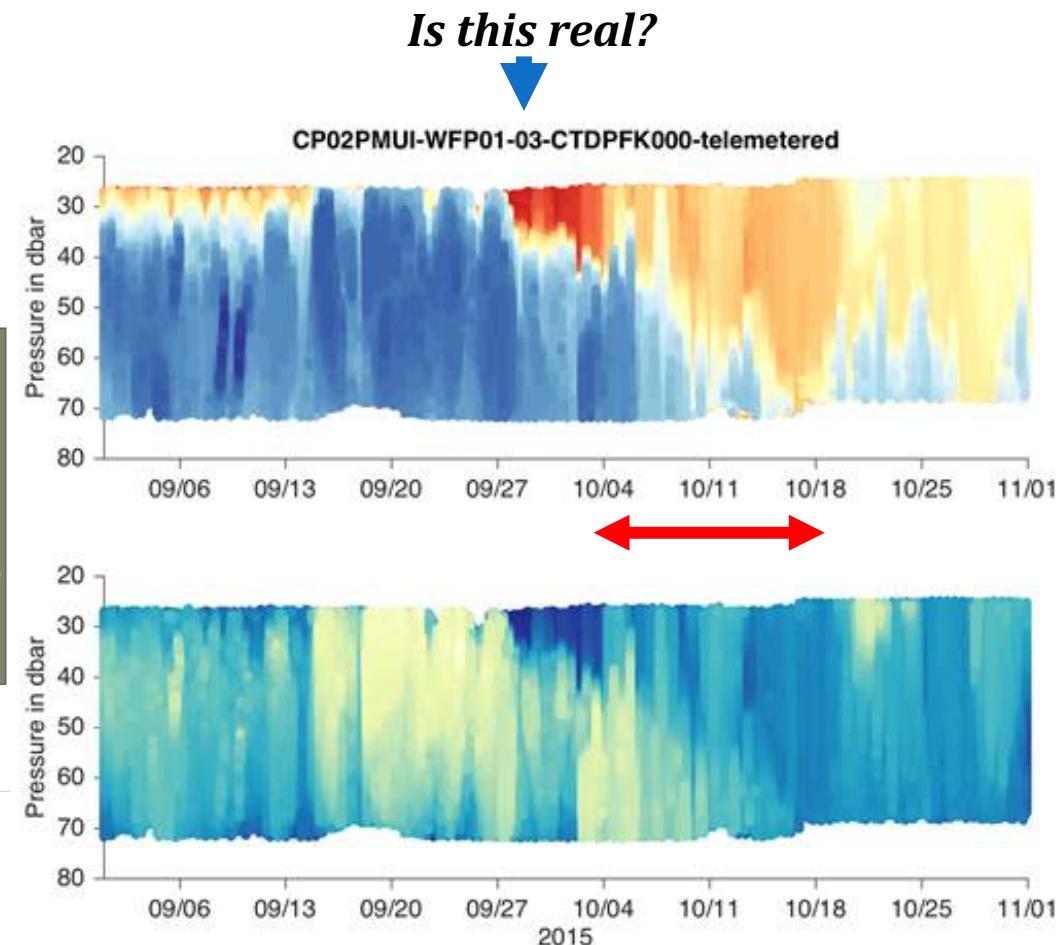
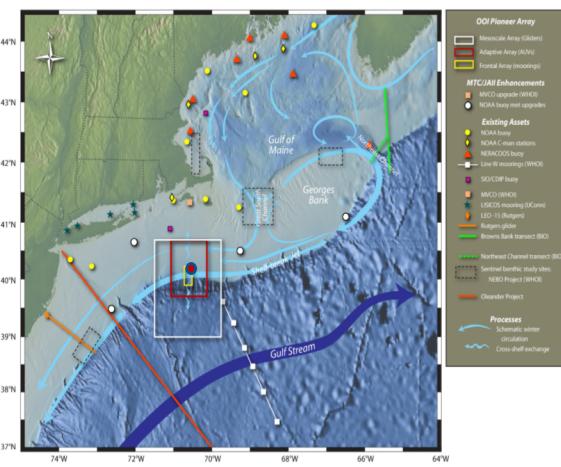
Northward Sea Water Velocity, m s⁻¹

Eastward Sea Water Velocity, m s⁻¹

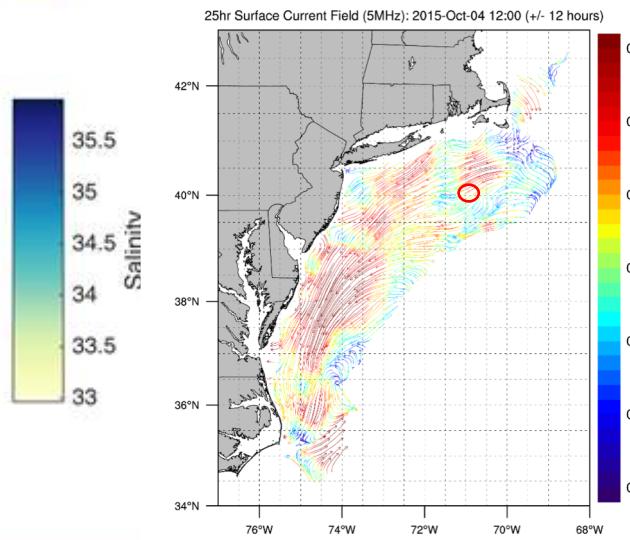


# Science Evaluation: Are ocean features encountered real? Outside local range

Taking advantage of all assets, even non-NSF, to assess data quality



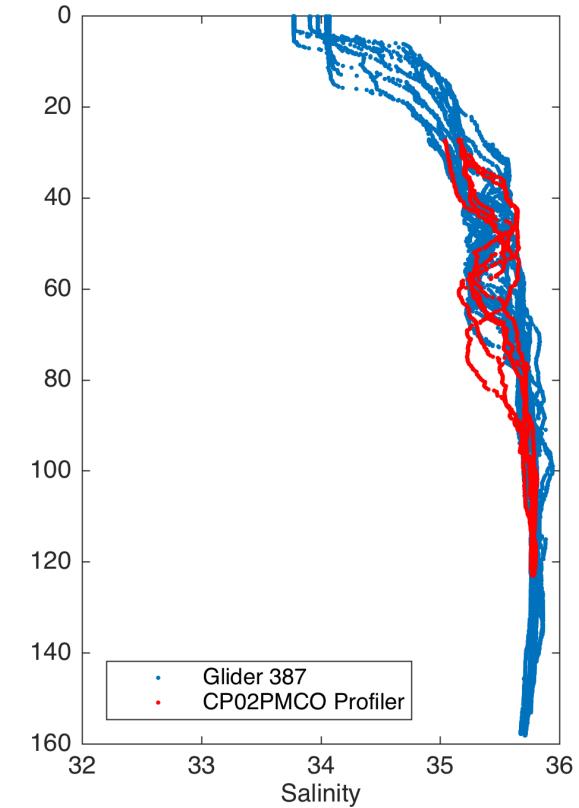
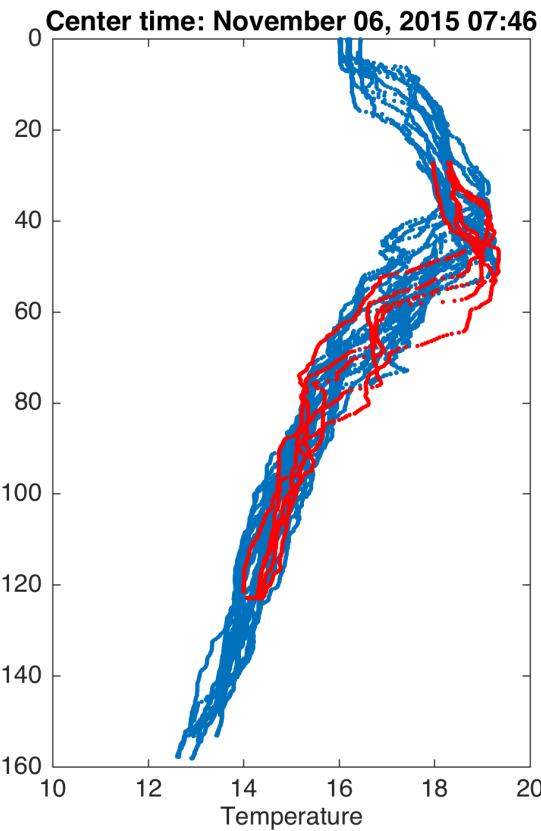
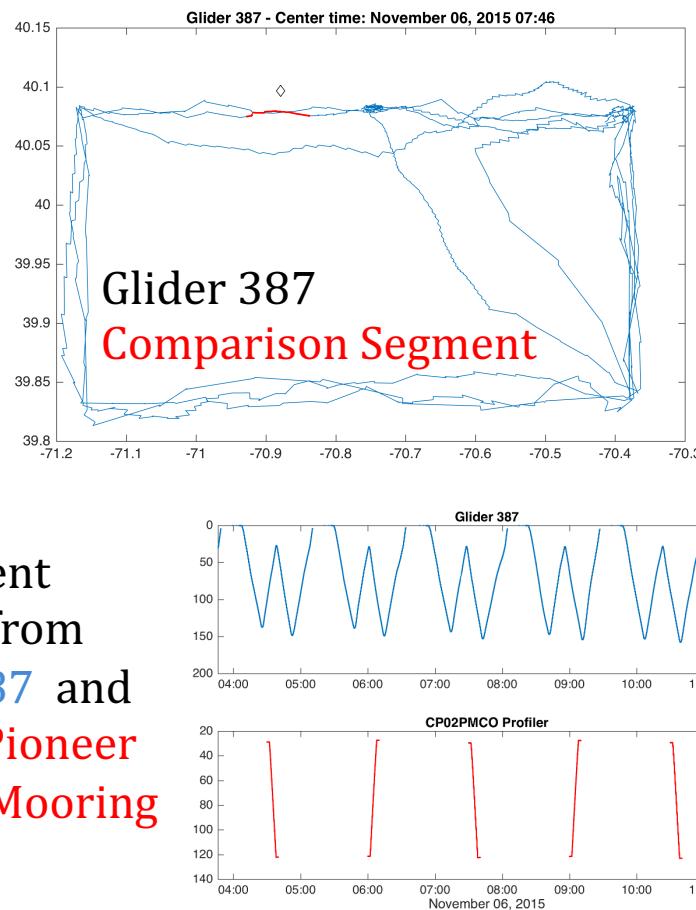
IOOS data shows presence of warm core ring



IOOS data shows ocean response to storm



# Vicarious Calibration - Comparisons enabled by ERDDAP

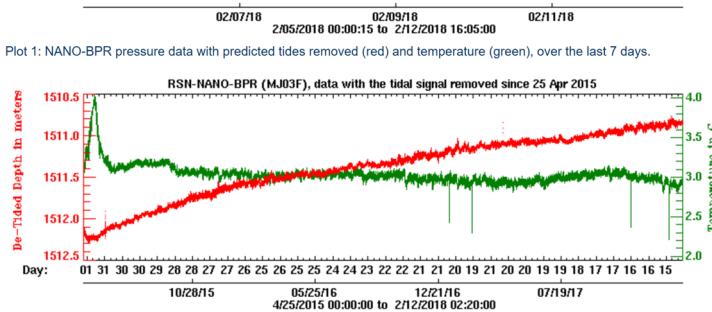
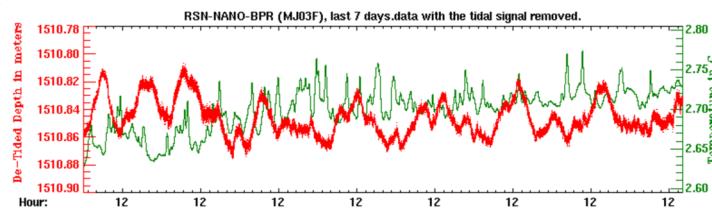


Co-located and concurrent Temperature and Salinity Profiles  
 Blue – Glider 387      Red – Coastal Pioneer Profiler Mooring



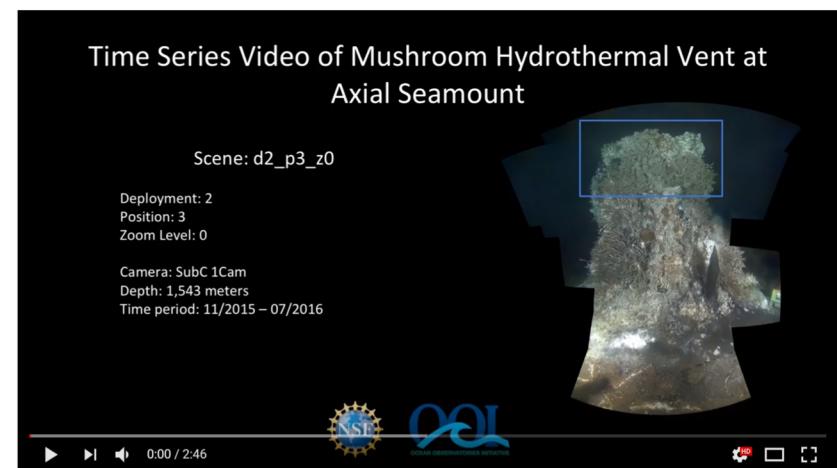
# Community Datasets & Tools

<http://oceanobservatories.org/community-tools/>



De-tided Bottom Pressure  
Recorder (BOTPT) data  
*W. Chadwick (OSU/CIMRS)*

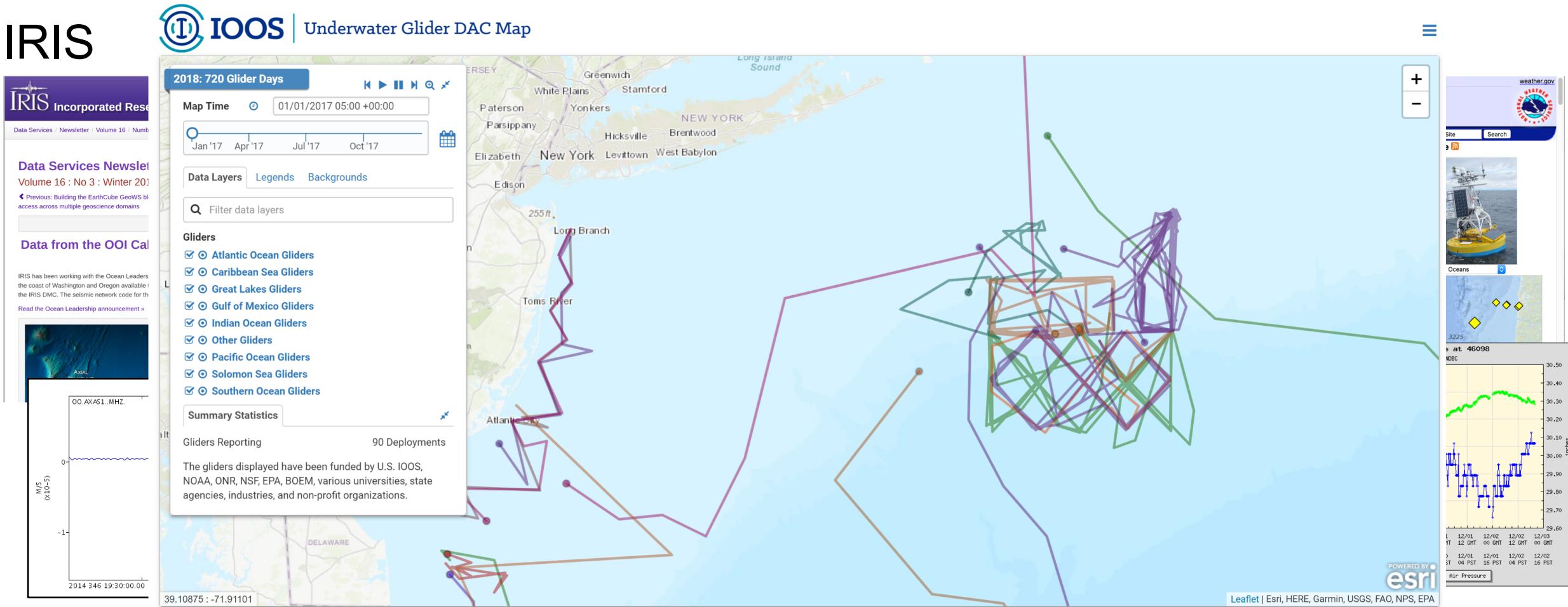
Student-collated biology catalog  
from Axial Seamount, using  
ROV and HD video records  
*D. Kelley (UW) & L. Sautter  
(Charleston)*



Time-lapse videos of vent dynamics at Axial Seamount, using HD Video data  
*A. Marburg (UW/APL), T. Crone (LDEO) & F. Knuth (Rutgers)*



# Partnerships



# CI Priorities (Transition to 2.0)

- Data product creation: Hydrophone data, Osmosampler, pCO<sub>2</sub> Flux, VADCP, interpolation issues
- Data discovery: navigation of data catalog, missing pressure data, image galleries, stream name consistency
- User experience: login fixes, interpolated plots, data download issues (missing status, provenance errors), status indicators and dashboard, addition of multiple-stream download
- Quality control: data displayed outside deployment time range, QC lookup values, fill value support, improved annotations



# Overview

1. Data Flow & Products
2. Data Review Procedures
3. Periodic Reviews & Documentation
4. Next Steps
5. Conclusions



# Conclusions

1. A large amount of high quality data continues to be collected
2. Data review is the primary focus, given maturation of the system and control over data ingestion
3. Data team accelerating review via development of specialized tools
4. Short-term goals for improving data quality and delivery (medium- and long-term goals are now recommendations for OOI 2.0)
5. OOI is providing a curated, consistent data system delivering data and metadata to the community



# Rutgers Data Team



Mike Crowley



Mike Vardaro



Sage  
Lichtenwalner



John Kerfoot



Leila  
Belabassi



Mike Smith



Friedrich  
Knuth



Lori Garzio

# Questions?

- OOI Main Web site: <http://oceanobservatories.org>
- Data Portal: <http://ooinet.oceanobservatories.org>

Help Desk: [help@oceanobservatories.org](mailto:help@oceanobservatories.org)

Mike Vardaro, Data Manager, OOI CI Data Team  
[vardaro@marine.rutgers.edu](mailto:vardaro@marine.rutgers.edu)

Mike Crowley, Program Manager, OOI CI Data Team  
[crowley@marine.rutgers.edu](mailto:crowley@marine.rutgers.edu)

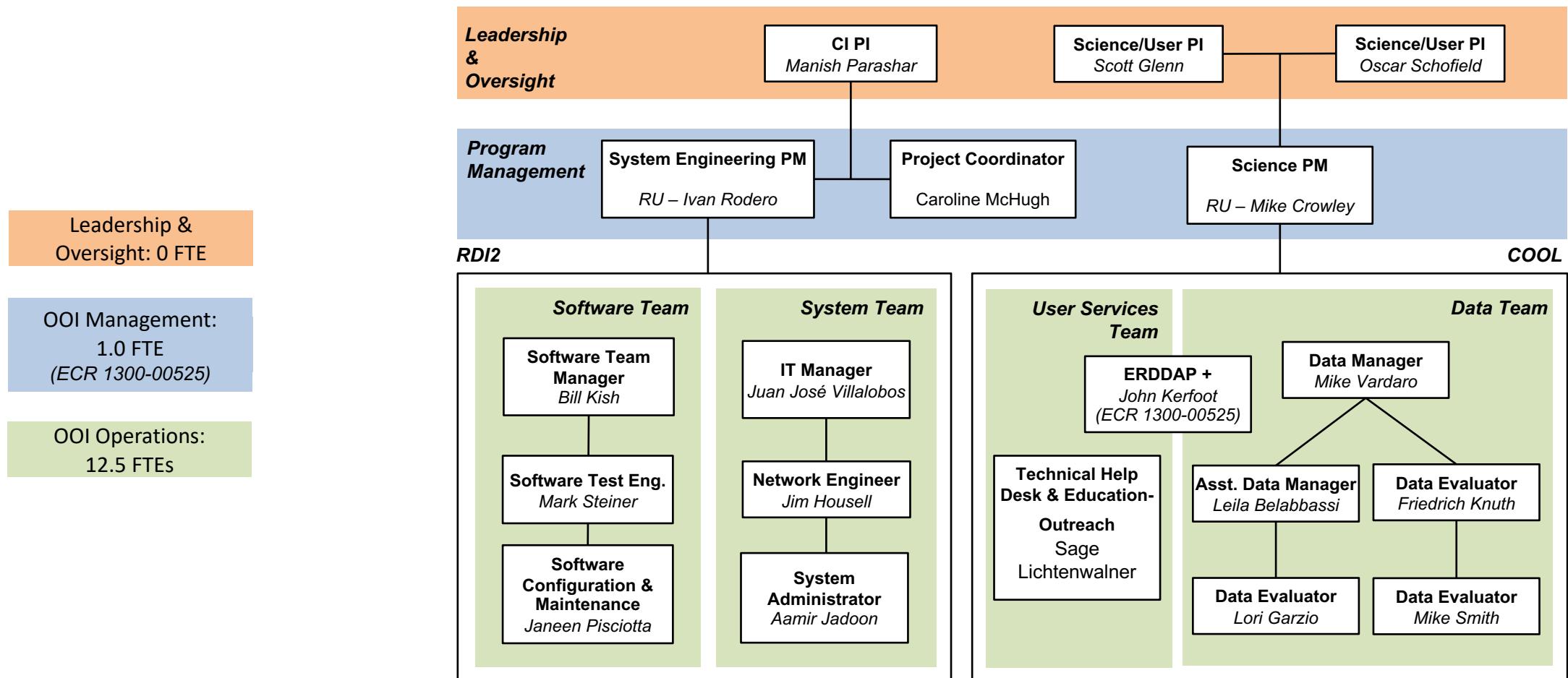
**Acknowledgements:** NSF, COL, Rutgers University, University of Washington, WHOI, Oregon State University, Raytheon



# Supplemental Slides



# Rutgers CI Team



# First in Class Reviews: Jan-Aug 2016

- One example of each data stream (ingestion completed by Systems team)
- Review of 1207 (467 science) streams completed in August 2016
- Tested parsers, algorithms, ingestion, asset management and data product creation

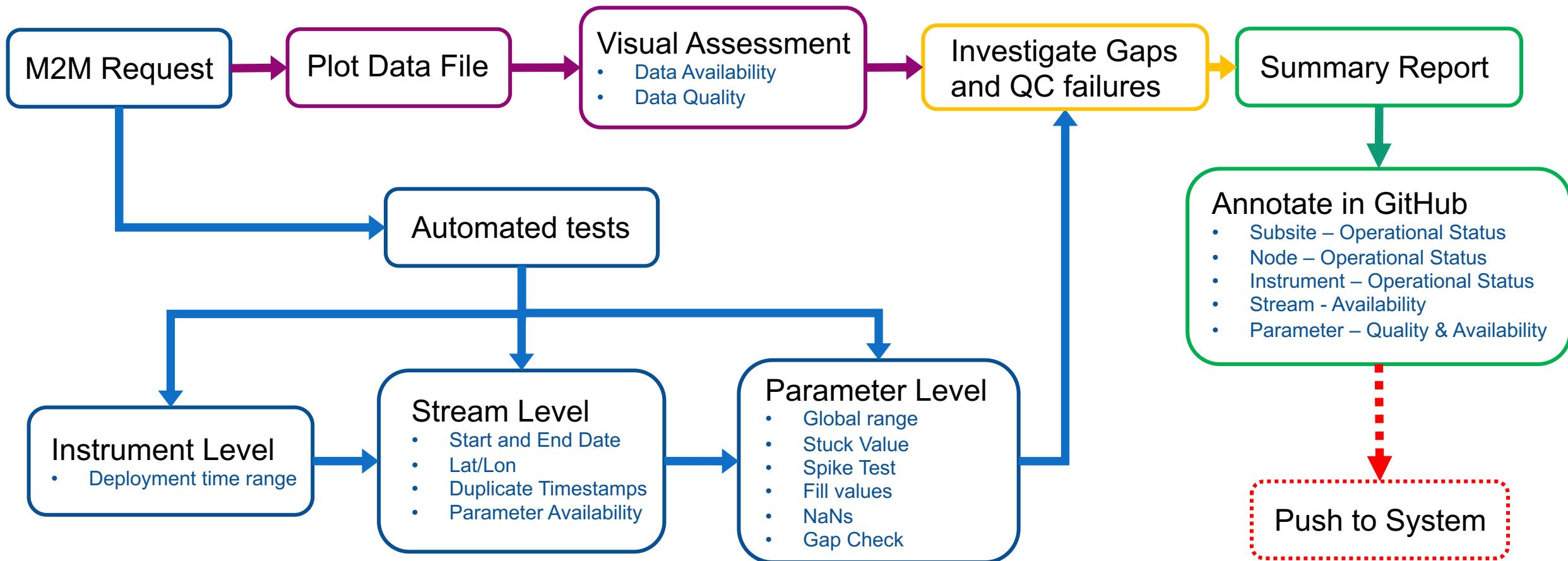
	WBS	Task Name	% Completions	Duration	Start	Finish
299	1.5.1	▪ Data Ingestion	62%	130 days	Wed 1/20/16	Tue 7/19/16
300	1.5.1.1	▪ First In Class for Cassandra Team	88%	62 days	Mon 1/25/16	Tue 4/19/16
301	1.5.1.1.1	▪ Pioneer Coastal Glider, CP05MOAS-GL388	100%	8 days	Mon 1/25/16	Wed 2/3/16
308	1.5.1.1.2	▪ Pioneer Central Inshore Profiler Mooring, CP02PMCI	100%	8 days	Thu 1/28/16	Mon 2/8/16
315	1.5.1.1.3	▪ Endurance OR Offshore Surface Mooring - CE09OSSM	100%	30 days	Fri 1/29/16	Thu 3/10/16
322	1.5.1.1.4	▪ Cabled Slope Base Low Power Jbox - RS01SLBS-LJ01A	84%	46 days	Thu 1/28/16	Thu 3/31/16
329	1.5.1.1.5	▪ Cabled Slope Base Shallow Profiler Mooring - RS01SBPS	100%	43 days	Thu 1/28/16	Mon 3/28/16
336	1.5.1.1.6	▪ Cabled Slope Base Deep Profiler Mooring - SRS01SBPD	100%	31 days	Fri 1/29/16	Fri 3/11/16
343	1.5.1.1.7	▪ Irminger Sea Apex Surface Mooring, GI01SUMO	100%	36.95 days	Tue 2/2/16	Wed 3/23/16
350	1.5.1.1.8	▪ Irminger Sea Apex Profiler Mooring (GI02HYPM)	100%	33.5 days	Tue 2/2/16	Fri 3/18/16
357	1.5.1.1.9	▪ Irminger Sea Flanking Subsurface Mooring A (GI03FLMA)	63%	33 days	Wed 2/3/16	Fri 3/18/16
364	1.5.1.1.10	▪ Irminger Global Open Ocean Glider (GIOSMOAS-GL)	100%	19.33 days	Wed 2/3/16	Tue 3/1/16
371	1.5.1.1.11	▪ Irminger Global Profiling Gliders (GI05MOAS-PG)	100%	31.5 days	Thu 2/4/16	Fri 3/18/16
378	1.5.1.1.12	▪ Coastal Endurance OR Inshore Surface Piercing Profiler Mooring (CE01ISSP)	100%	30.5 days	Fri 2/5/16	Fri 3/18/16
385	1.5.1.1.13	▪ Coastal Endurance OR offshore BEP - CE04OSBP	31%	30 days	Wed 2/10/16	Tue 3/22/16
392	1.5.1.1.14	▪ Cabled Seafloor Instruments	0%	12 days	Mon 4/4/16	Tue 4/19/16



	WBS	Task Name	% Completions	Duration	Start	Finish
418	1.5.2	▪ Data Verification & Validation	28%	255 days	Wed 2/10/16	Tue 1/31/17
419	1.5.2.1	▪ First in Class	39%	123 days	Tue 3/1/16	Thu 8/18/16
420	1.5.2.1.1	Pioneer Coastal Glider, CP05MOAS-GL388	95%	34 days	Tue 3/1/16	Fri 4/15/16
421	1.5.2.1.2	Endurance OR Offshore Surface Mooring - CE09OSSM	70%	30 days	Fri 3/4/16	Thu 4/14/16
422	1.5.2.1.3	Pioneer Upstream Inshore Profiler Mooring, CP02PMCI	42%	31.8 days	Fri 3/18/16	Mon 5/2/16
423	1.5.2.1.4	Cabled Slope Base Deep Profiler Mooring - RS01SBPI	0%	20 days	Tue 6/7/16	Mon 7/4/16
424	1.5.2.1.5	Cabled Slope Base Low Power Jbox - RS01SLBS-LJ01A	0%	27 days	Fri 4/29/16	Mon 6/6/16
425	1.5.2.1.6	Cabled Slope Base Shallow Profiler Mooring - RS01SI	75%	33 days	Tue 3/15/16	Thu 4/28/16
426	1.5.2.1.7	Irminger Sea Apex Profiler Mooring (GI02HYPM)	100%	20 days	Fri 3/25/16	Thu 4/21/16
427	1.5.2.1.8	Irminger Global Open Ocean Glider (GI05MOAS-GL)	0%	3 days	Fri 4/22/16	Tue 4/26/16
428	1.5.2.1.9	Irminger Sea Flanking Subsurface Mooring A (GI03FLMA)	0%	18 days	Wed 4/27/16	Fri 5/20/16
429	1.5.2.1.10	Irminger Global Profiling Gliders (GI05MOAS-PG)	0%	5 days	Mon 5/23/16	Fri 5/27/16
430	1.5.2.1.11	Irminger (Or other global) Sea Apex Surface Mooring	40%	74 days	Mon 3/14/16	Thu 6/23/16
431	1.5.2.1.12	Coastal Endurance OR Inshore Surface Piercing Profiler	0%	8 days	Fri 6/24/16	Tue 7/5/16
432	1.5.2.1.13	Coastal Endurance OR offshore BEP - CE04OSBP	0%	22 days	Fri 7/8/16	Mon 8/8/16
433	1.5.2.1.14	Cabled Axial Seamount Central Caldera Med Power J	0%	10 days	Fri 7/8/16	Thu 7/21/16
434	1.5.2.1.15	Cabled Seafloor Instruments	0%	20 days	Fri 7/22/16	Thu 8/18/16
435	1.5.2.1.16	AUVs	0%	6 days	Tue 8/9/16	Tue 8/16/16
436	1.5.2.2	▪ AGU Oceans Data Prep (THREDDS & GUI) - Reasonability	100%	30 days	Wed 2/10/16	Tue 3/22/16



# Rest in Class Data Review Workflow

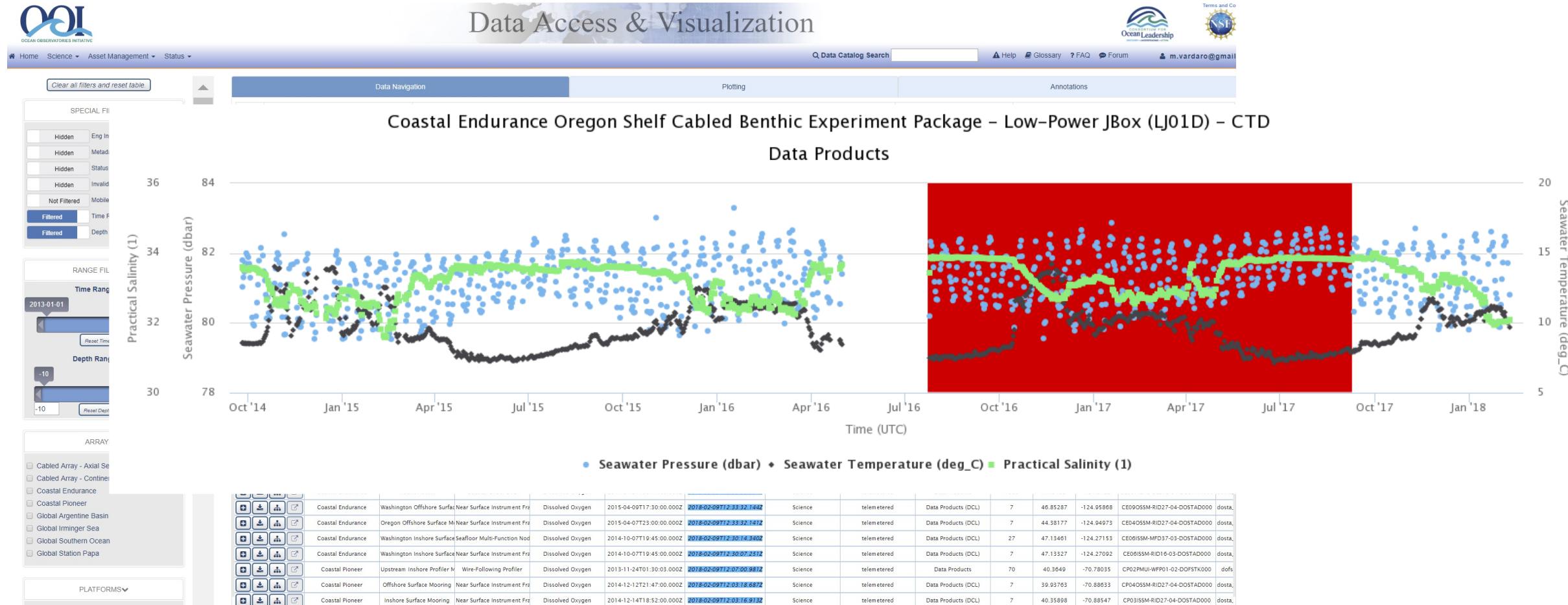


# QARTOD/OOI QC Comparison

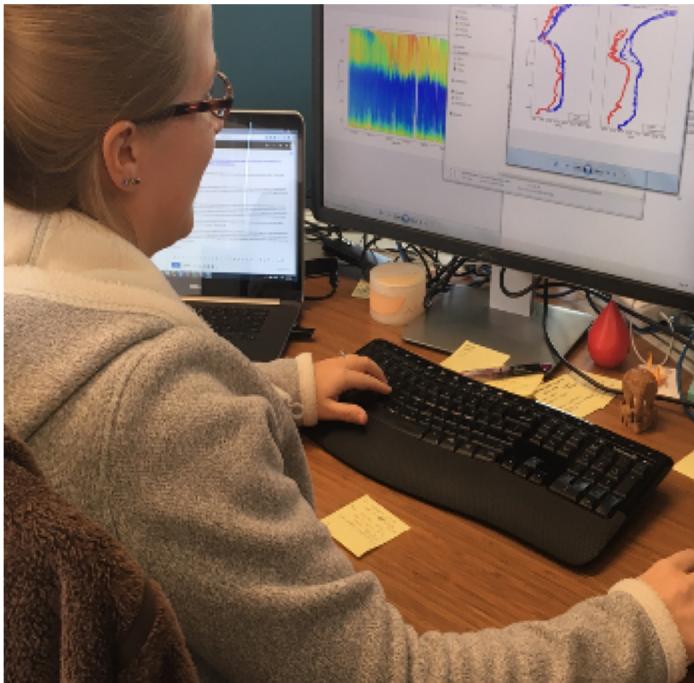
OOI Test	OOI Description	QARTOD Equivalent	QARTOD Recommendation (from manuals)	Notes
Global Range Test	Data are flagged unless they fall within valid world ocean ranges or instrument limits	Gross Range	Only considers manufacturer-defined sensor and calibration limits	Different tests, different names. <b>Currently operational.</b>
Spike Test	Deviation from mean compared to 2*N neighboring points	Spike	N=1, default threshold is based on the rate of change distribution from previous data sets	Roughly identical, same nomenclature. <b>Currently operational.</b>
Stuck Value Test	If 2 neighboring values differ by less than the resolution of the sensor for more than N repetitions, data are flagged	Stuck Sensor	Manual suggests 3 consecutive points for a stuck sensor suspect flag and 5 for a fail flag.	QARTOD manual suggestion may be too low for well-mixed portions of the water column. <b>Under evaluation.</b>
Local Range Test	Data are flagged unless they fall within locally valid site-specific or depth ranges. Interpolates thresholds between depth and season intervals	Local Range	Starts with constant limits for each depth/season interval	Roughly identical, same nomenclature. <b>OOI Local ranges are still being established.</b>
Gradient Test	If $d(\text{data})/d(t) >$ a set threshold, following points fail until one falls within limit (TOLDAT). First data point assumed good unless “good” starting data (STARTDAT) point is defined.	Rate of Change	QARTOD recommends two neighboring points and does not incorporate TOLDAT or STARTDAT values.	Different tests, different names. <b>Under Evaluation, not operational</b>
Trend Test	Data flagged as having trend if the SD of the residuals to a polynomial curve < original data, multiplied by a factor. Test for sensor drift.	N/A	No QARTOD equivalent	OOI only. <b>Under Evaluation, not operational</b>



# Quick Plots and Data Download

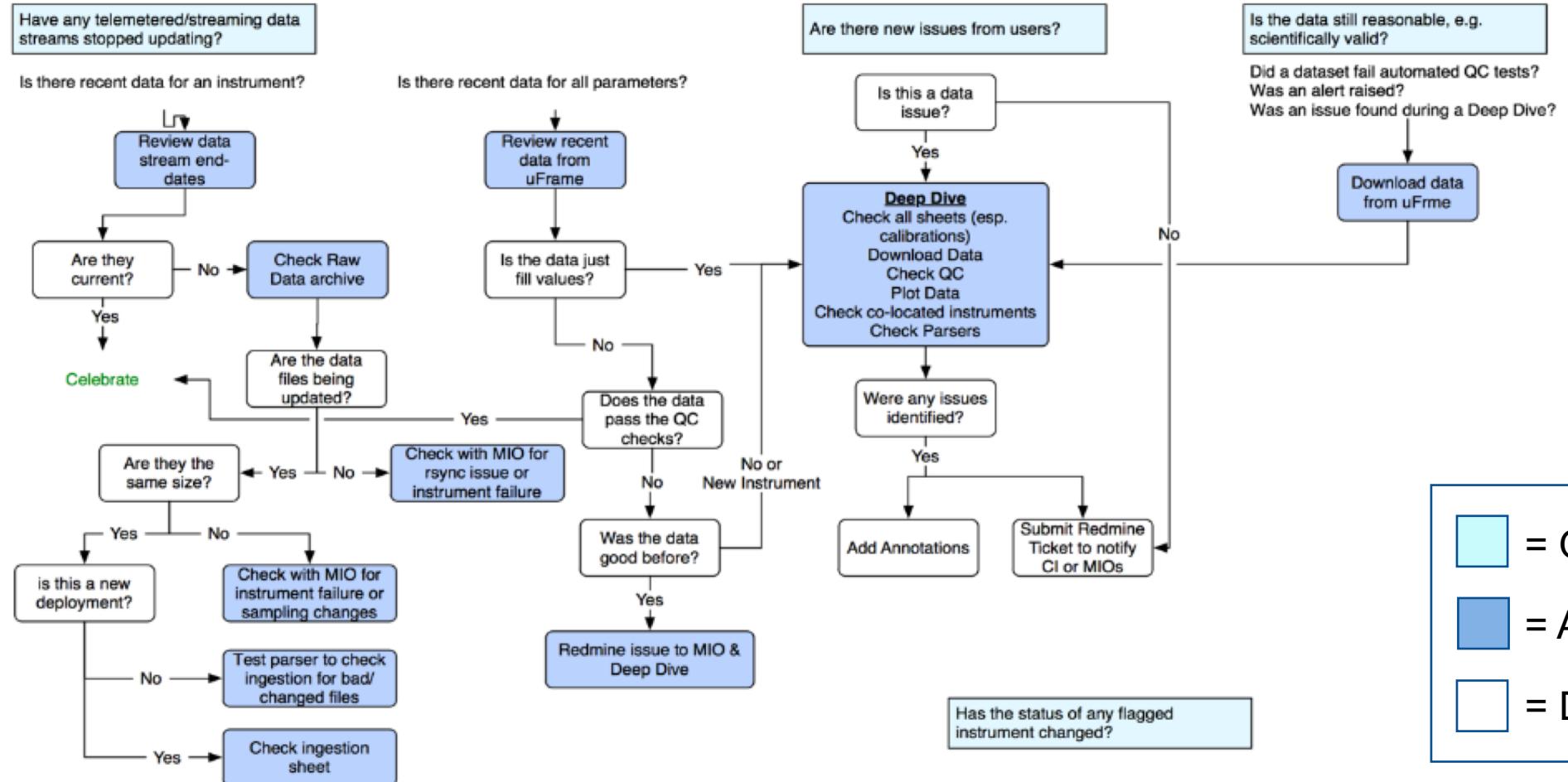


# Data Evaluation Daily Activities



- Review the end-to-end operational status of online instruments and investigate any outages (e.g. instrument, telemetry, parsing, or ingestion failures).
- Review the operational status of other data archives (raw, cruise, ERDDAP)
- Look into and resolve new system alerts
- Follow up on any issue requests from users (via Redmine)
- Add annotations to notify users of operational status changes

# Daily Review Workflow



- [Light Blue Box] = Questions
- [Dark Blue Box] = Actions
- [White Box] = Decision points



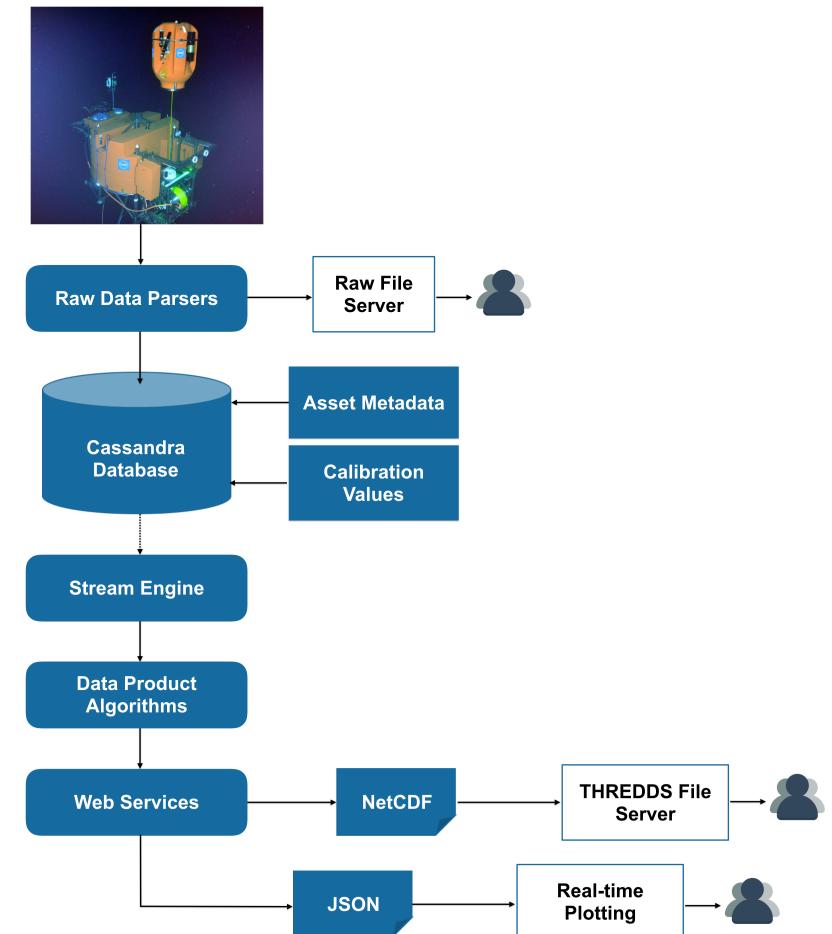
# QC Challenges & Solutions

- Local range values need statistical analysis of environmental data for each platform
  - ***Need to work with SMEs to analyze and apply ranges and test algorithm***
- Trend test may not work as designed, because it requires the system to compare data prior to the user request date – ***analysis ongoing***
- Gradient test is complicated to apply, requires 2D dataset – ***analysis ongoing***
- Spike test is currently very simple - ***needs tweaking to avoid false positives/negatives (especially in biological data) and to work with certain data types***
- Not all QC algorithms apply to all data products – ***ongoing review with SMEs***
- The QC algorithms do NOT trigger alerts in the system - ***Alerts/alarms only trigger when new data is telemetered/streamed***
  - ***Can set alerts on L1/L2 data streams based on Global/Local range values***



# Machine-to-Machine RESTful API

- Raw data, metadata, & calibration values extracted on demand, algorithms applied
- NetCDF data/metadata always the full time range requested, assembled once data products are created (asynchronous)
- Synchronous JSON requests decimated down to a maximum of 20,000 data points
- JSON response is instantaneous, making it useful for real-time plotting applications



# OOI ERDDAP

## Current ERDDAP Status

- 1655 streams available
- Telemetered, currently deployed systems
- Data available via uFrame

## Future Updates

- Additional telemetered data sets updated as deployed
- Cabled Array data
- Recovered data (requires additional storage)
- Multidimensional dataset issues

ERDDAP > List of All Datasets

Pick a Dataset

1394 matching datasets, listed in alphabetical order. View page: 1 (current) 2 .

Grid	DAP Data	Sub-set	Table Data	Make Graph	W Data	M Data	Source Files	Title	Summary	FGDC, ISO, Metadata	Back-ground Info	RSS	E mail	Institution	Dataset ID
set	data	graph						* The List of All Active Datasets in this ERDDAP *	?	M	background	?	OOI		allDatasets
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame 3-Wavelength Fluorometer float_dcl_instrument - Deployment 0007 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-02-FLORTD000-flort_dcl_instrument-telemetered-deployment0007-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame CTD ctldb_p_cdef_dcl_instrument - Deployment 0007 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-03-CTDBPC000-ctldb_p_cdef_dcl_instrument-telemetered-deployment0007-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_cpu_optime - Deployment 0007 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_cpu_optime-telemetered-deployment0007-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_cpu_uptime - Deployment 0008 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_cpu_uptime-telemetered-deployment0008-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_dlog_mngr - Deployment 0007 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_dlog_mngr-telemetered-deployment0007-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_dlog_mngr - Deployment 0008 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_dlog_mngr-telemetered-deployment0008-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_dlog_status - Deployment 0007 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_dlog_status-telemetered-deployment0007-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_dlog_status - Deployment 0008 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_dlog_status-telemetered-deployment0008-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_error - Deployment 0007 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_error-telemetered-deployment0007-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_error - Deployment 0008 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_error-telemetered-deployment0008-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_gps - Deployment 0007 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_gps-telemetered-deployment0007-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_gps - Deployment 0008 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_gps-telemetered-deployment0008-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_msq_counts - Deployment 0007 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_msq_counts-telemetered-deployment0007-tabledap	
								Coastal Endurance Oregon Inshore Surface Mooring Near Surface Instrument Frame Data Concentrator Logger (DCL) cg_dcl_eng_dcl_msq_counts - Deployment 0008 (telemetered)	?	E I M	background	?	OOI	CE01ISSM-RD16-00-DCLENG000-cg_dcl_eng_dcl_msq_counts-telemetered-deployment0008-tabledap	

<https://erddap-uncabled.oceanobservatories.org/uncabled/erddap/index.html>



# Current “Rest-in-Class” Reviews

## Process:

- Check all deployments for presence & absence of all parameters
- Check science parameters for reasonableness
- Problem? Deep dive, report in Redmine, track, give feedback, check fixes, create annotations in QC Database

## Challenge:

- Automated tools, Redmine questions, Cal sheets, raw data repository, modify ingest CSVs, testing UI fixes
- Upload and ingestion of data
- Delivery and archiving of Cruise Data
- Quality Assurance vs. Quality Control

## Expediting the Solution:

- Populate QC database to automatically check for presence/absence, gaps > 1 day, NaNs, negative values

1. Asset Management  
(MIOs & Data Team)
  - Complete?
  - Correct?

2. Data Delivery & Ingestion  
(MIOs, Systems, Data Team)
  - Includes Cruise Data

3. Data Review
  - Availability
  - Quality

4. Investigate Gaps and QC failures

5. Communicate Issues (Annotation)



# Options for Data Review Acceleration

Option	Positive	Negative
MIO Operations Log at Rutgers	<ul style="list-style-type: none"> <li>Centralized log reduces time spent investigating issues</li> <li>All issues entered consistently</li> </ul>	<ul style="list-style-type: none"> <li>Takes time to maintain</li> <li>Some development time</li> </ul>
Speed up ingestion	<ul style="list-style-type: none"> <li>Fewer gaps to investigate</li> </ul>	<ul style="list-style-type: none"> <li>Currently requires FTE</li> <li>Automated process not yet delivered</li> </ul>
Data Team works only on RIC	<ul style="list-style-type: none"> <li>More data reviewed faster</li> </ul>	<ul style="list-style-type: none"> <li>No new data in system</li> <li>No bug investigation</li> <li>No QA testing</li> </ul>
Limit reviewed time period or stream type	<ul style="list-style-type: none"> <li>Data reviewed slightly faster, at high level</li> </ul>	<ul style="list-style-type: none"> <li>Review enhanced by looking at multiple deployments and trends</li> <li>Slows down future reviews</li> </ul>
Limit thoroughness of reviews	<ul style="list-style-type: none"> <li>Data reviewed faster, at high level</li> </ul>	<ul style="list-style-type: none"> <li>Unclear why gaps exist</li> <li>Quality issues not fully annotated</li> <li>Slows down future reviews</li> <li>Limits crowdsourcing options</li> </ul>
Crowdsourcing (enlist volunteer SMEs)	<ul style="list-style-type: none"> <li>Removes subset of datasets from review queue</li> <li>Assistance with complex data that requires expertise</li> </ul>	<ul style="list-style-type: none"> <li>Focus on specific interest, not whole of OOI</li> <li>Steep learning curve for advanced use of system (and knowledge of known issues)</li> <li>Pathway to triage and incorporate feedback</li> </ul>
Add employees or Data Assembly Center (DAC)	<ul style="list-style-type: none"> <li>Data reviewed faster, in depth</li> <li>Support for expert analysis</li> </ul>	<ul style="list-style-type: none"> <li>Requires additional funding</li> <li>Setup and maintenance time</li> </ul>

# Periodic Data Team Activities

- Meet with MIOs to discuss operational issues and data quality
- Instrument, stream, parameter and deployment completeness
- Conduct deep dives on datasets to review availability and quality
- Review & annotate full deployment data to assess data quality
- Develop new scripts, plotting tools, and quality checks
- Produce reports on the availability and quality of datasets
- Review appropriateness of QC flags
- Ensure asset, deployment, calibration, and ingestion configurations have been updated, and reports posted following every cruise
- Prototype and test new user interface and visualization features

