OOI CI Data Team Update
September 2017

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

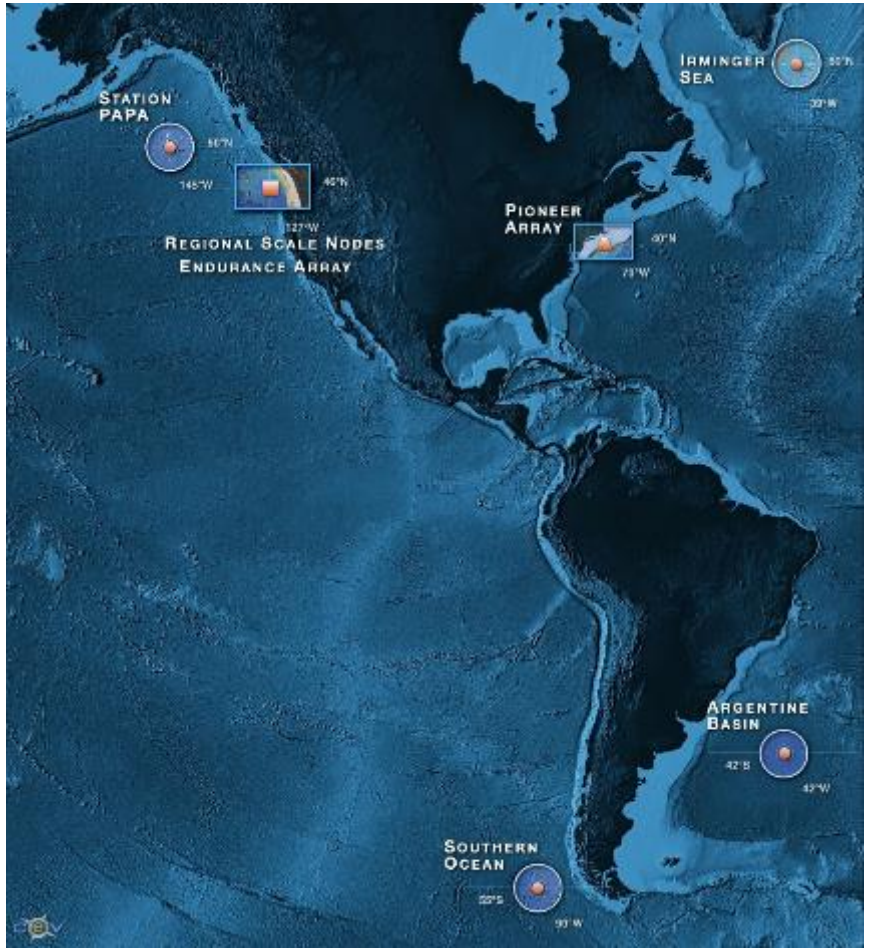
Rutgers University, New Brunswick, NJ

Overview

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

   a) A large amount of high value data is being collected
   b) Data ingestion & review is our primary focus
   c) Accelerating data review via development of specialized tools
   d) Short-, medium-, and long-term goals to improve data quality
   e) OOI is providing a curated, consistent data system that is delivering data and metadata to the community

OOI By the Numbers



7 Arrays

57 Stable Platforms
Moorings, Profilers, Nodes

32 Mobile Assets
Gliders, AUVs

1227 Instruments (857 deployed)

>2500 Science Data Products

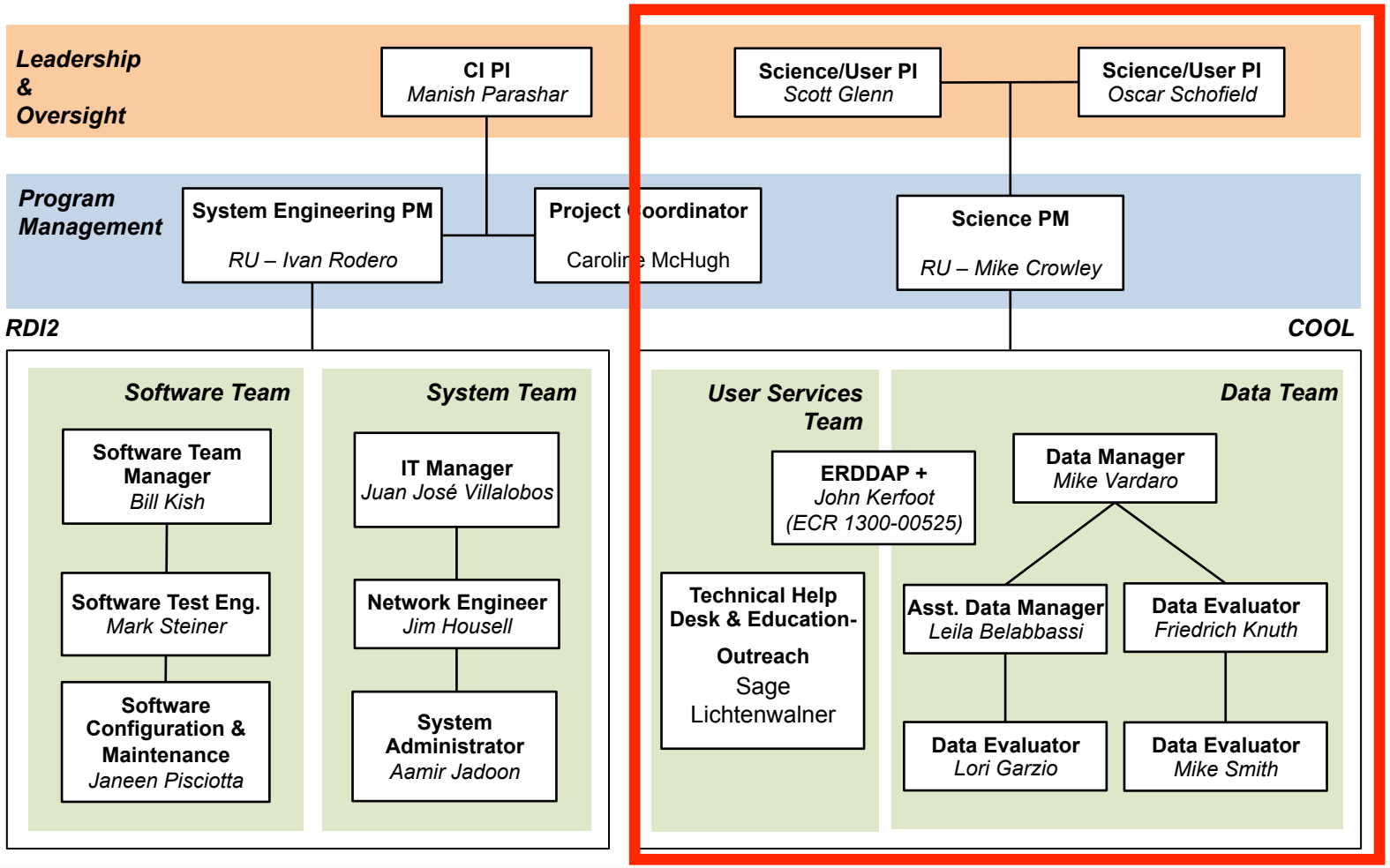
>100K Science/Engineering Data Products

Rutgers CI Team

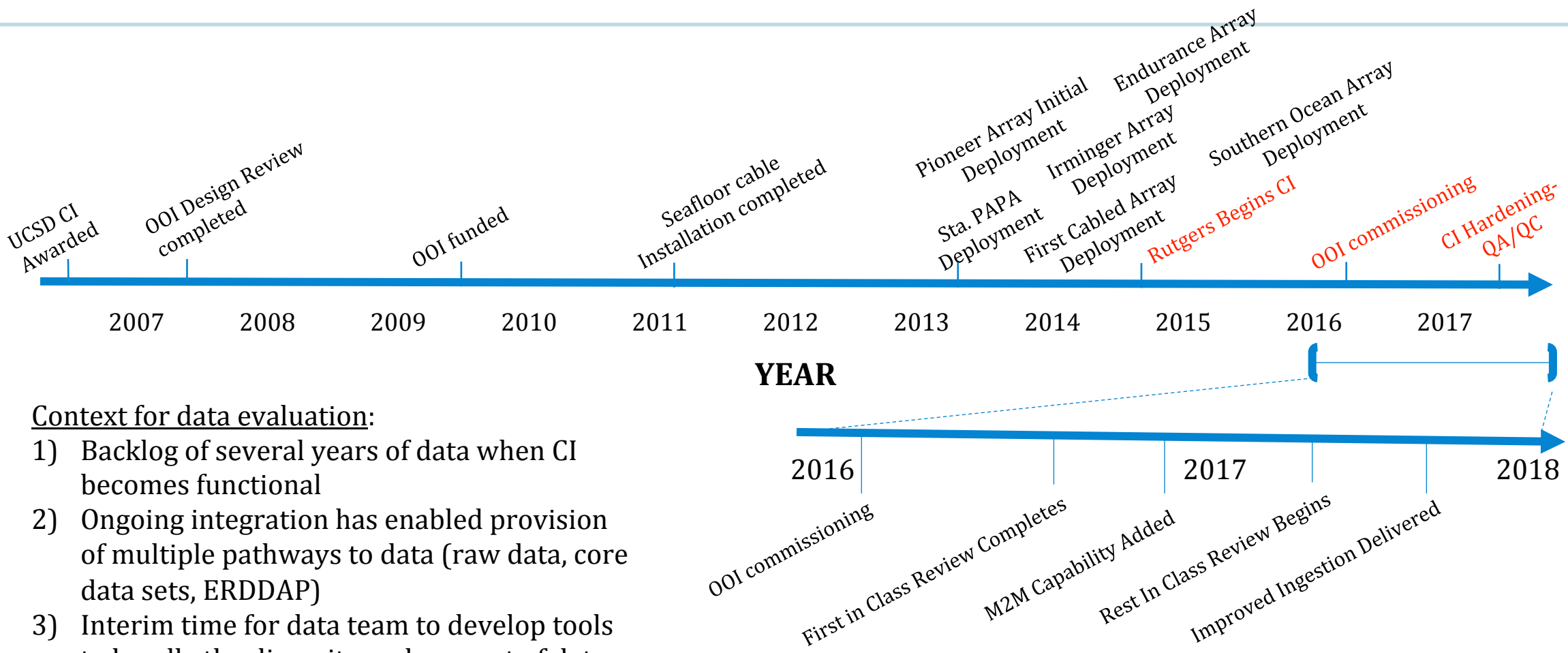
Leadership & Oversight: 0 FTE

OOI Management: 1.0 FTE *(ECR 1300-00525)*

OOI Operations: 12.5 FTEs



Timeline Context



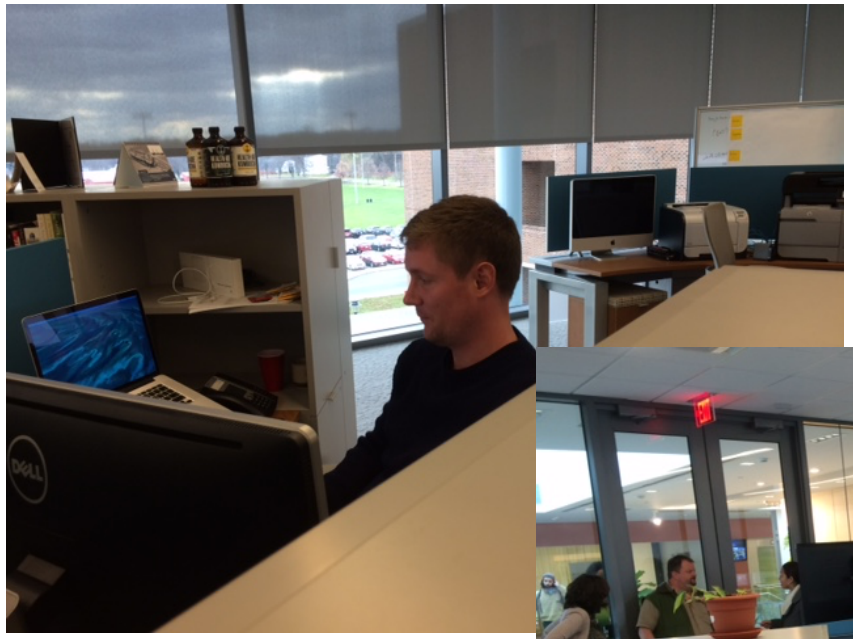
<u>Context for data evaluation</u>:

1) Backlog of several years of data when CI becomes functional
2) Ongoing integration has enabled provision of multiple pathways to data (raw data, core data sets, ERDDAP)
3) Interim time for data team to develop tools to handle the diversity and amount of data

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. Identify availability and quality issues and ensure they are resolved
5. Communicate known data issues with end users
6. Report operational statistics on data availability, data quality, and issue resolution

Day in the life of an OOI data evaluator

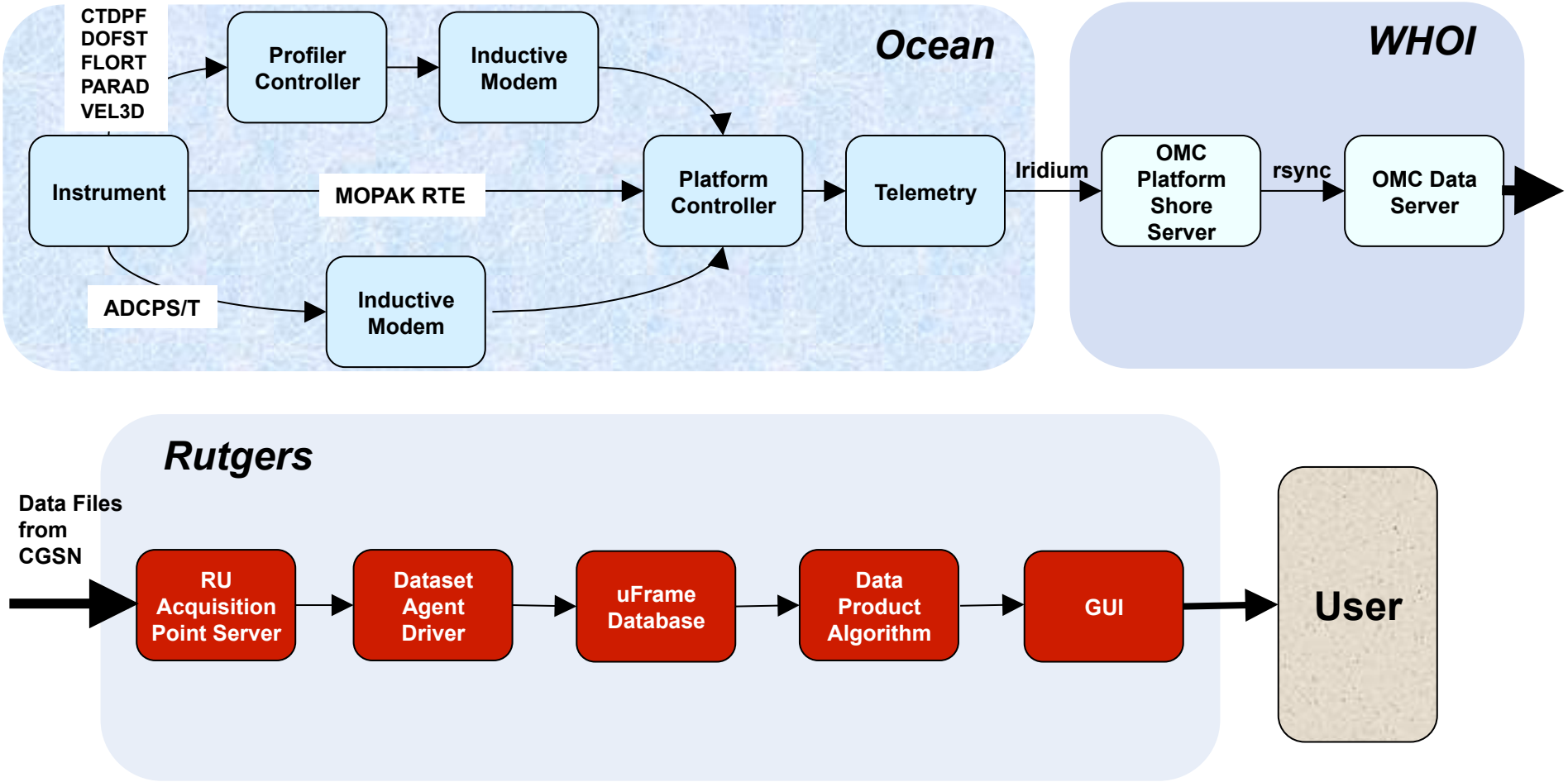


- Evaluator is assigned a specific array and works closely with the MIOs
- Range of expertise (biology, physics, geophysics)
- Data community including collocated non-OOI data experts
- Development of open-access tools to visualize and synthesize the data
- Developing inputs for automated QC
- Quick looks and deep data dives, updating OOINet asset sheets
- Interactions within team, with MIOs, with users & students, and specific SMEs
- Full-time effort is required

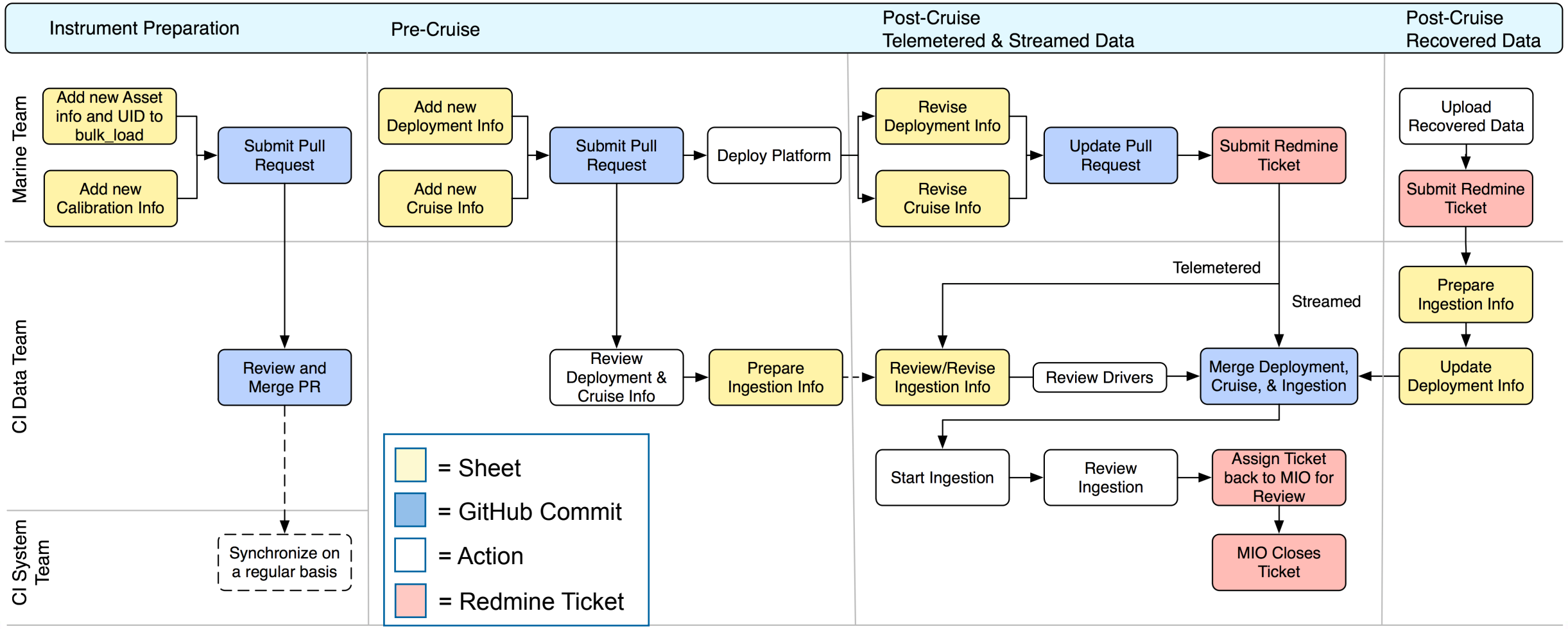
Overview

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

Data Flow Example: Pioneer Profiler



Current Data Processing Flow



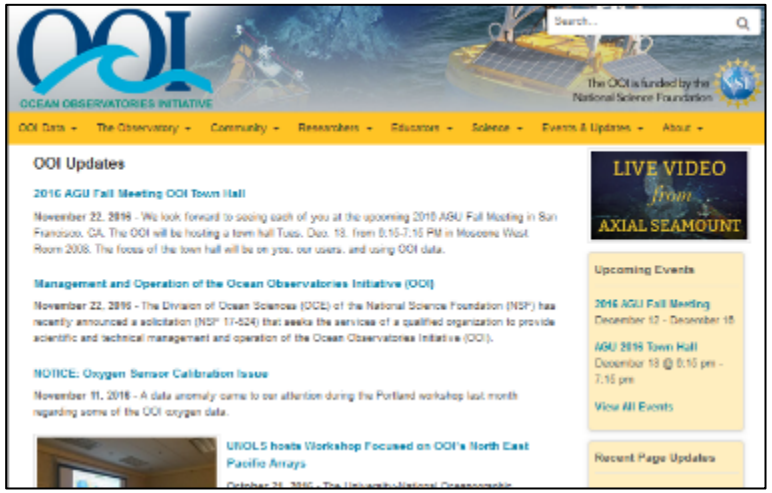
Data Types

- Telemetered Data
  - Data received through a transmission medium over distance (e.g. surface buoy to satellite, glider to satellite, acoustic modem); may be decimated
- Recovered Data
  - Data downloaded directly from a recovered instrument or data logger after the instrument has been recovered.
- Streamed Data
  - Data received via transmission over electro-optical cable. Streaming data are provided at full temporal resolution and near-real time.
- Shipboard Data
  - Shipboard data and water samples collected during OOI expeditions.
- Metadata
  - Info about the data record (e.g., time & location of collection, unique source & record description identifier, instrument serial #, etc.). OOI metadata follows the CF1.6 standard, with additional types and fields specific to OOI as necessary.

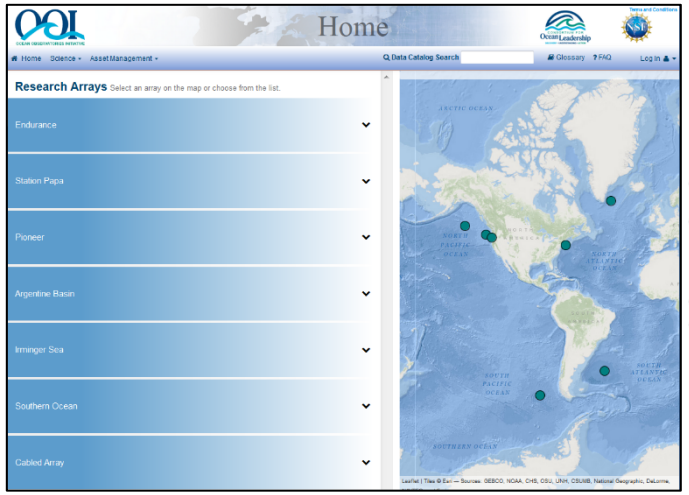
OOI Data Product Levels

- **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
  - QC may be applied at this level, utilizing simple automated techniques or human inspection
  - 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
  - Products may come from the same or from separate instruments
  - Data from all relevant instruments will be provided during download
  - Example: SBE-37IM Density and Salinity

OOI: Web Portals

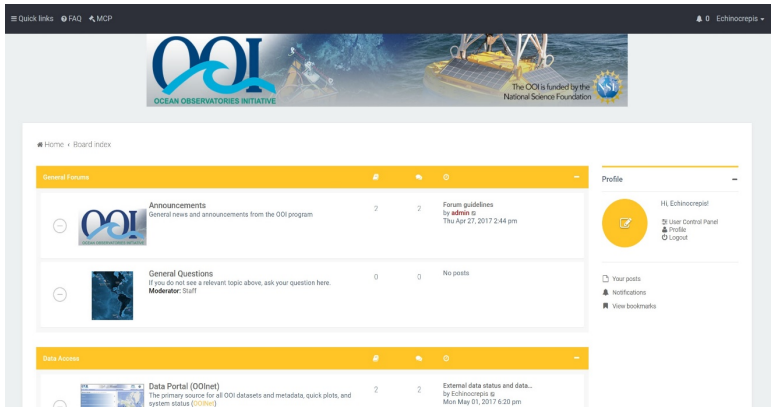


Main

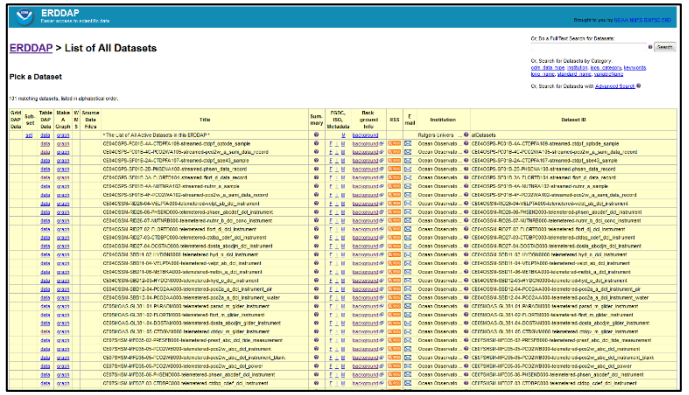


OOINet

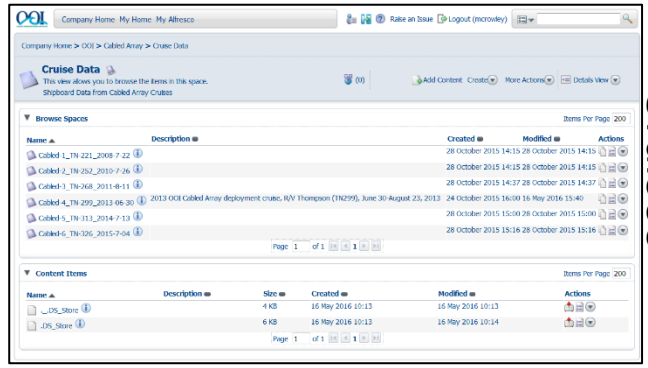
Education



Forum



ERDDAP



Cruises

Overview

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

First in Class Reviews: Jan-Aug 2016

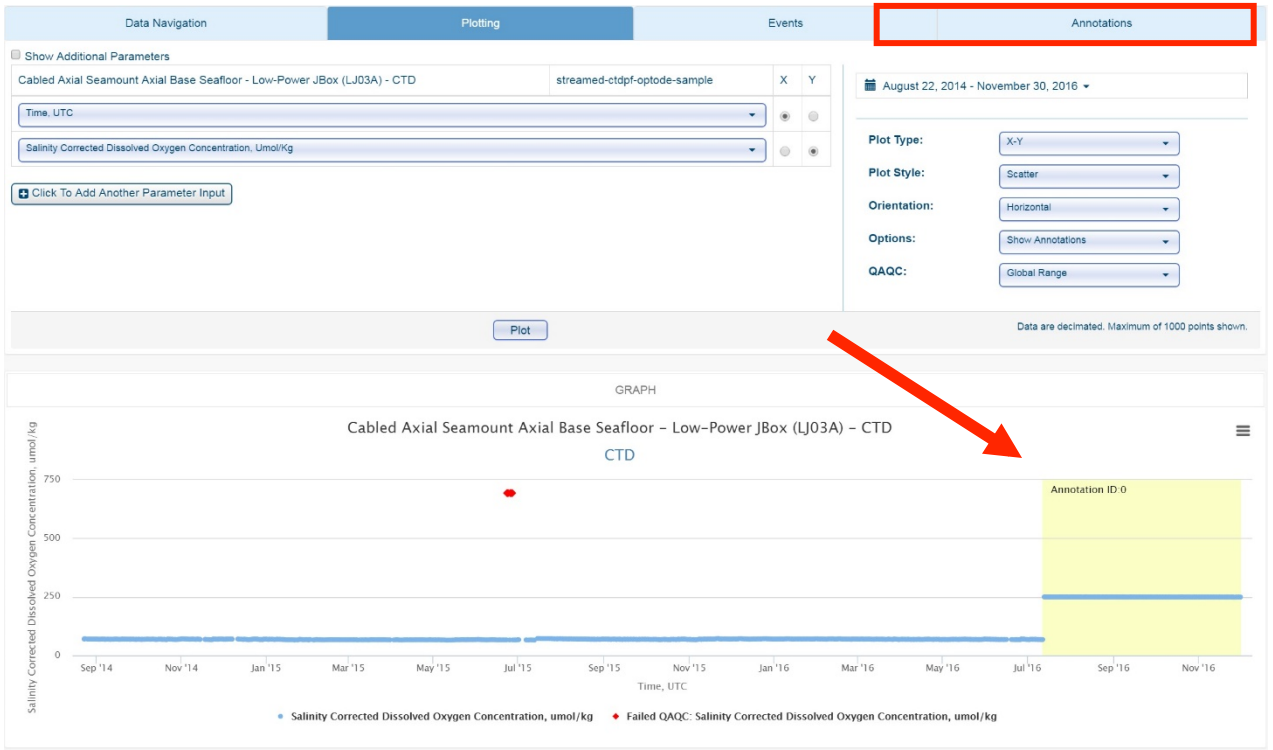
- One example of each data stream (ingestion completed by Systems team)
- Review of 1207 (467 science) streams completed in August 2016
- Tested parsers, algorithms, ingestion, asset management and data product creation

	WBS	Task Name	% Complete	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 (GIO5MOAS-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	% Complete	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, CP02PM	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 (GIO5MOAS-GL)	0%	3 days	Fri 4/22/16	Tue 4/26/16
428	1.5.2.1.9	Irminger Sea Flanking Subsurface Mooring A (GI03FI	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 Profil	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**

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

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?

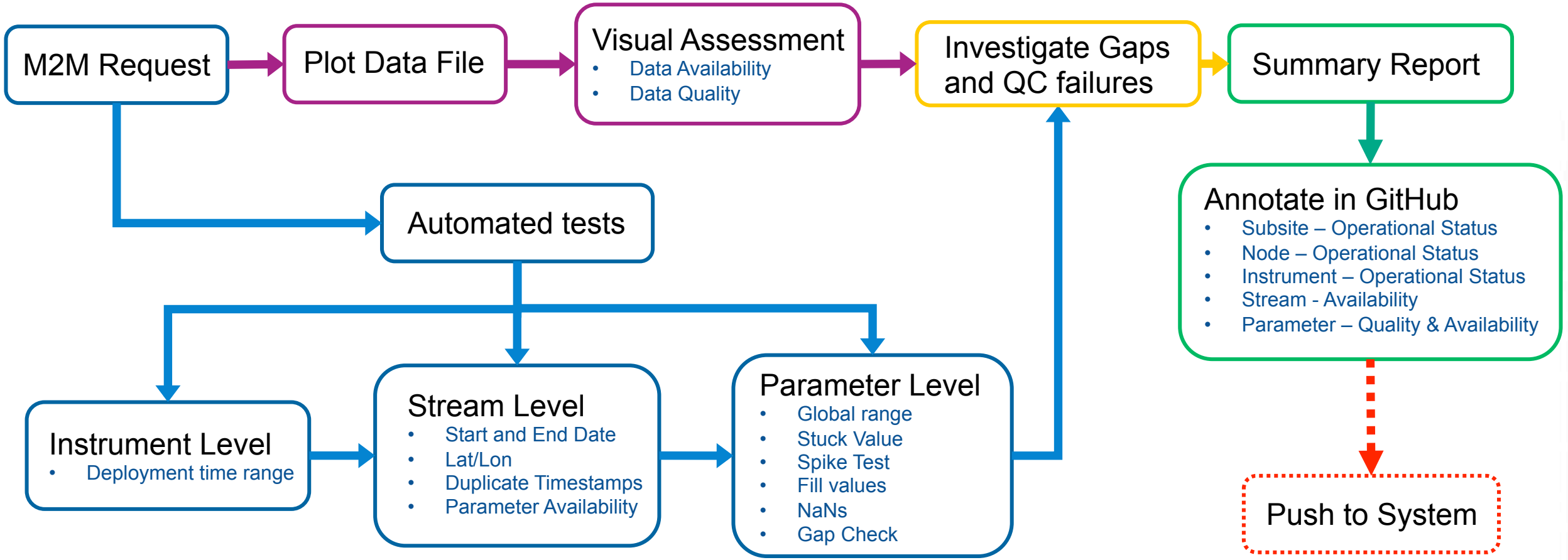
2. Data Delivery & Ingestion
   (MIOs, Systems, Data Team)
   - Includes Cruise Data

3. Data Review
   - Availability
   - Quality

4. Investigate Gaps and QC failures

5. Communicate Issues (Annotation)

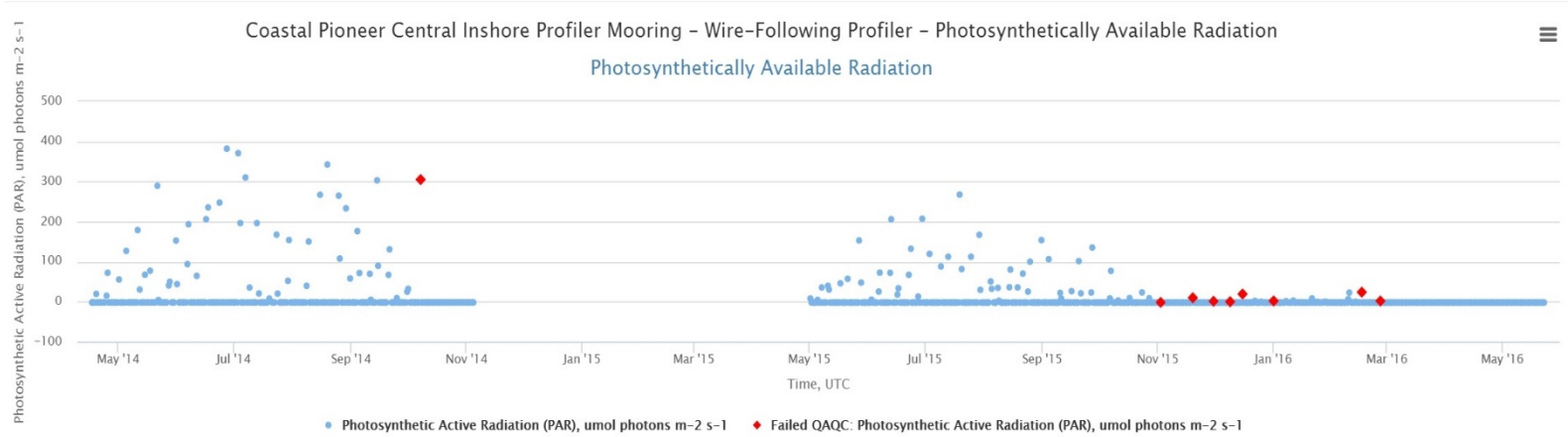
Rest in Class Data Review Workflow



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)



Rest in Class Data Status Categories

Status	Description	QARTOD Code	QARTOD Description	Color
NOT_OPERATIONAL	Instrument not functional (no data expected)		Not operational	
NOT_AVAILABLE	Instrument functional, data lost in transmission	9	Missing data	
PENDING_INGEST	Instrument functional, data exists, Awaiting ingest			
NOT_EVALUATED	Instrument functional, data exists, Awaiting evaluation	2	Not evaluated, not available or unknown	
SUSPECT	Instrument functional, data exists and either failed a QC test or does not reflect environmental conditions	3	Questionable/suspect	
FAIL	Instrument functional, data exists but is known to be bad due to known instrument or calibration error	4	Bad	
PASS	Instrument functional, data exists, passed QC tests, is complete and looks reasonable	1	Good	
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			

Data Availability

- 87.5% of all actively deployed platforms are providing all available data from the most recent deployments (1 Endurance glider and 3 Global platforms need updating)
- 75.3 Gb of processed data delivered in the last year
- 41.9 Tb of raw data delivered in the last year
- Recovered data and backlogged telemetered data are currently being ingested into the system sequentially
- Aiming to present full timeline “heat map” display by end of November early December 2017

Ingestion

- Testing purge/ingest for complete deployments to generate estimated ingest schedule for all data sets

Dependencies

- Delivery of deployed and recovered data to the raw data server at Rutgers: mostly up to date (latest check is ongoing).
- Ingest currently requires a follow-up query to determine that ingestion has completed
- Playback functionality required for purge/reingestion of cabled data (fill gaps and improve provenance): PRIORITY
- Purge/reingestion of uncabled profilers can take 24 hours to complete. Surface moorings may take 1.5 days, gliders 2 days, depending on number of deployments and amount of recovered data = 197 days for 113 uncabled platforms.
- Need configuration changes to allow multiple simultaneous ingestions to run more quickly. This will allow the 113 ingestions to occur in a compressed time frame by multiple team members.
- Need modification to the ingestion delay that eliminated the race condition issue, without which all telemetered data will be 24-48 hours delayed (not a blocker, but it does prevent real-time delivery and data status alerts)
- Purge functionality is still limited to purging an entire instrument (by reference designator) for all deployments and all delivery methods. Improving purge to allow purge by stream and time range will speed up the process.

Review

Current Status:

- Data review interrupted while the team focuses on ingest. Previously run reviews resulted in annotations that largely deal with unexplained data gaps, which we hope will be filled in by the purge and reingestion process.

Dependencies

- Completion of ingestion of recovered data sets that the data team is currently performing. Without ingestion, the sources of data gaps will not be able to be verified, requiring redundant checks.
- Data review can take 2-3 days to run per instrument (possibly less, depending on data volume and optimization of routines), and results in a report that can be used to enter annotations.
- Need to collate operational & cruise information from MIOs in order to compare to data availability/quality assessments and annotate properly

Annotation

Current Status:

- Annotations from previous data reviews logged within Sage Database (ooi.visualocean.net/) and uFrame (only the most urgent notifications were entered manually). The Sage Database is the more comprehensive, and ingest of this group of annotations into uFrame is the focus of this topic. Any additional annotations will require data quality review to proceed.

Dependencies

- Data review is required to create a full list of annotations for each data set, using automated routines, and visual inspection of plots.
- An API to push annotations at the platform, node, instrument, and stream level exists, but the manual entry process via the UI only currently allows stream-level annotations.
- Once review is complete for an instrument, annotations can be pushed to the uFrame system using the API.
- MIO operators can also enter annotations via the UI, if they are granted permissions and training, but there may be conflict between annotations entered via API and manual entry.

ERDDAP

Current Status:

- There are 846 streams available on the production server at https://erddap-uncabled.oceanobservatories.org/uncabled/erddap/info/index.html?page=1&itemsPerPage=1000

Dependencies:

- Login, edit and write access to production system:
  - Systems team must push the software updates as we do not have access to production.
  - Timely responses from systems team on requests for info through Redmine
  - Developer requires direct access to production to edit existing ERDDAP XML Files
- Some active tickets need resolution to remove caveats and inconsistencies in the data sets, including #12544 (Variable naming conventions for external L1/L2 parameters) and #10745 (NetCDF format: use of coordinate attribute and associated issues)
- As data streams get split and parameters merged (DOSTA/CTD, ADCP, etc.) ERDDAP developer needs to be informed, so that templates and datasets can be updated.

Demos

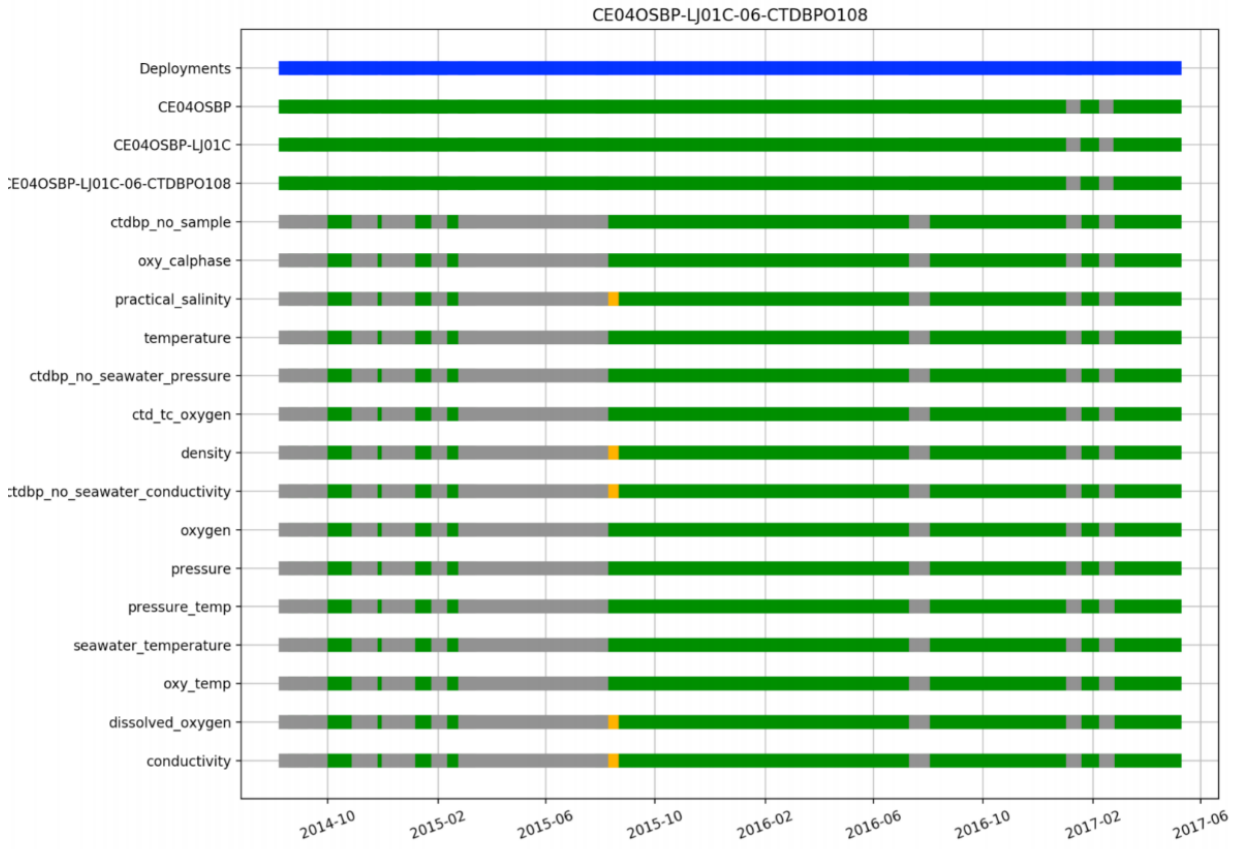
1. Main OOI website: oceanobservatories.org
2. Data Portal: ooinet.oceanobservatories.org
   a) Navigation
   b) Data catalog
   c) Quick Plotting options
   d) Data download
3. OOI ERDDAP server
4. M2M API and real-time plotting

Overview

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

Reviews and Reporting

Quality Timeline

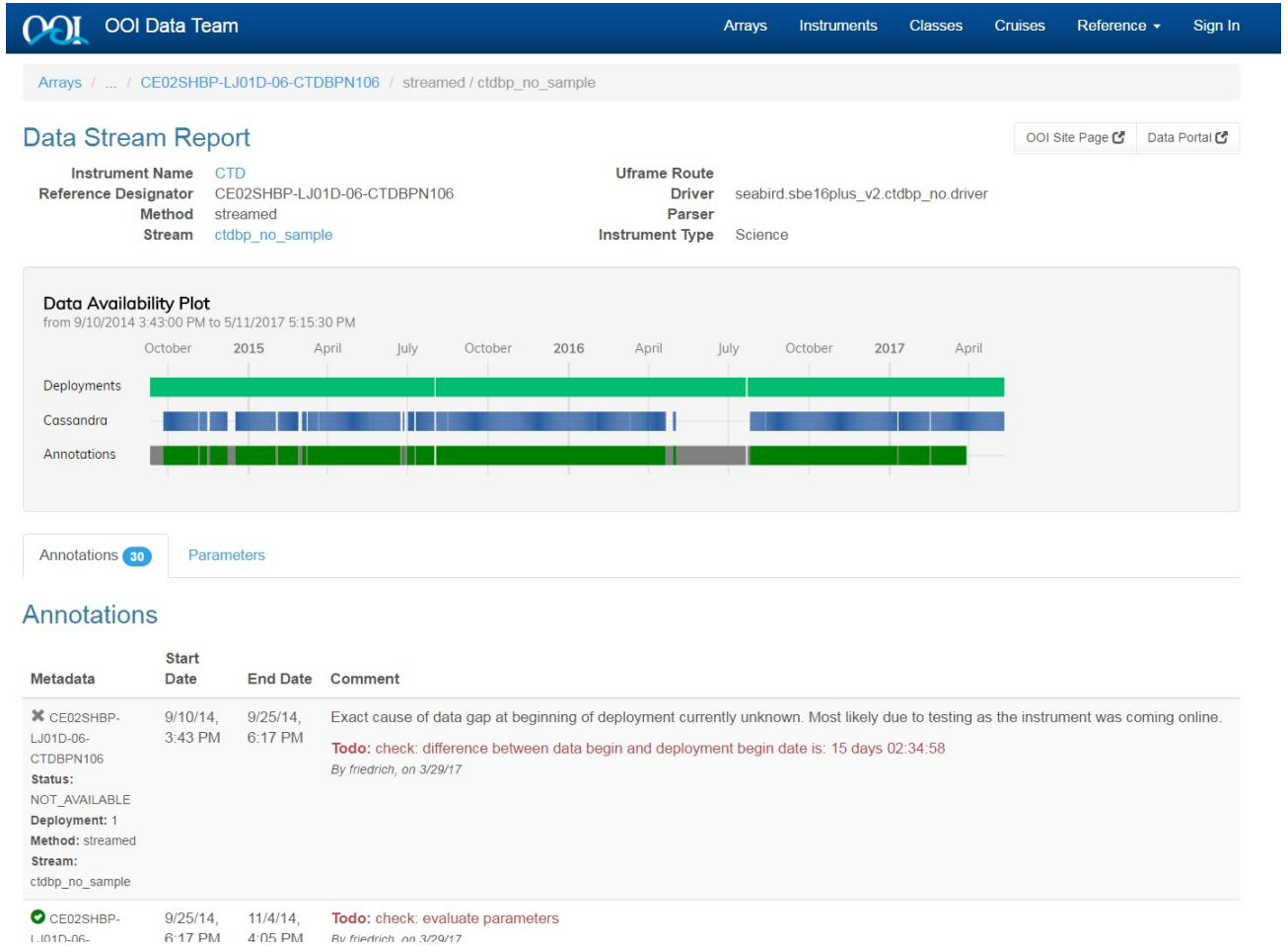


Annotation Text

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		2016-03-10T23:06:00Z	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	Corvalis 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

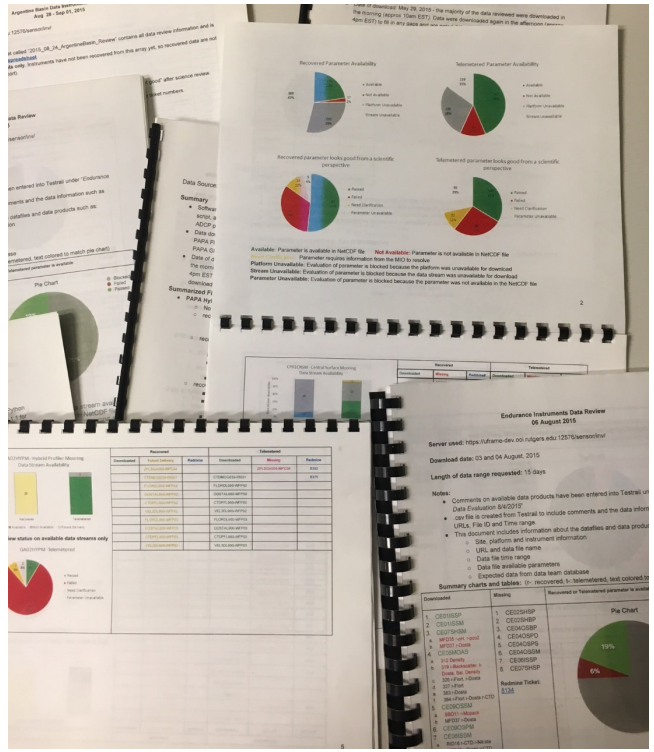
QC Database Tool

- Used for reference & statistics
- Includes status information, as well as a cruise data checklist
- Includes testing/review capability
- Annotation options
- http://ooi.visualocean.net

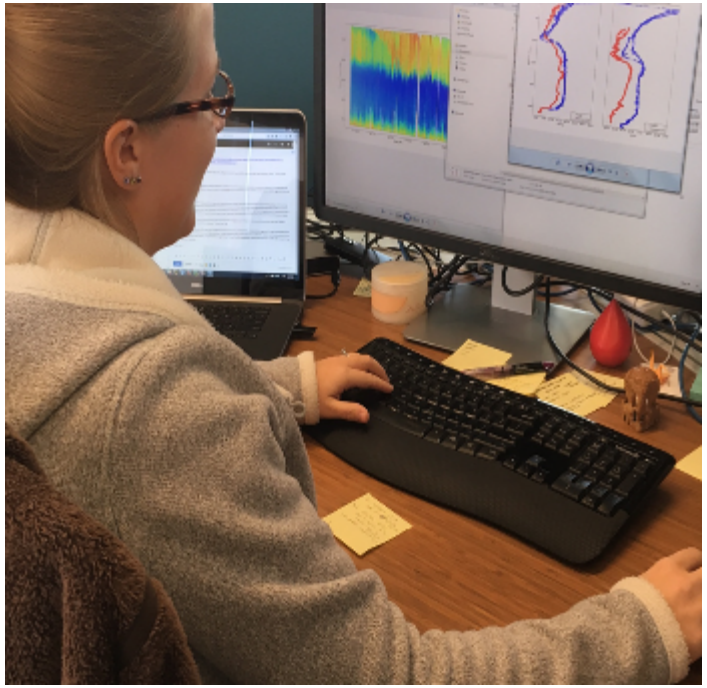


Deliverables

- 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
- Forum statistics (TBD)

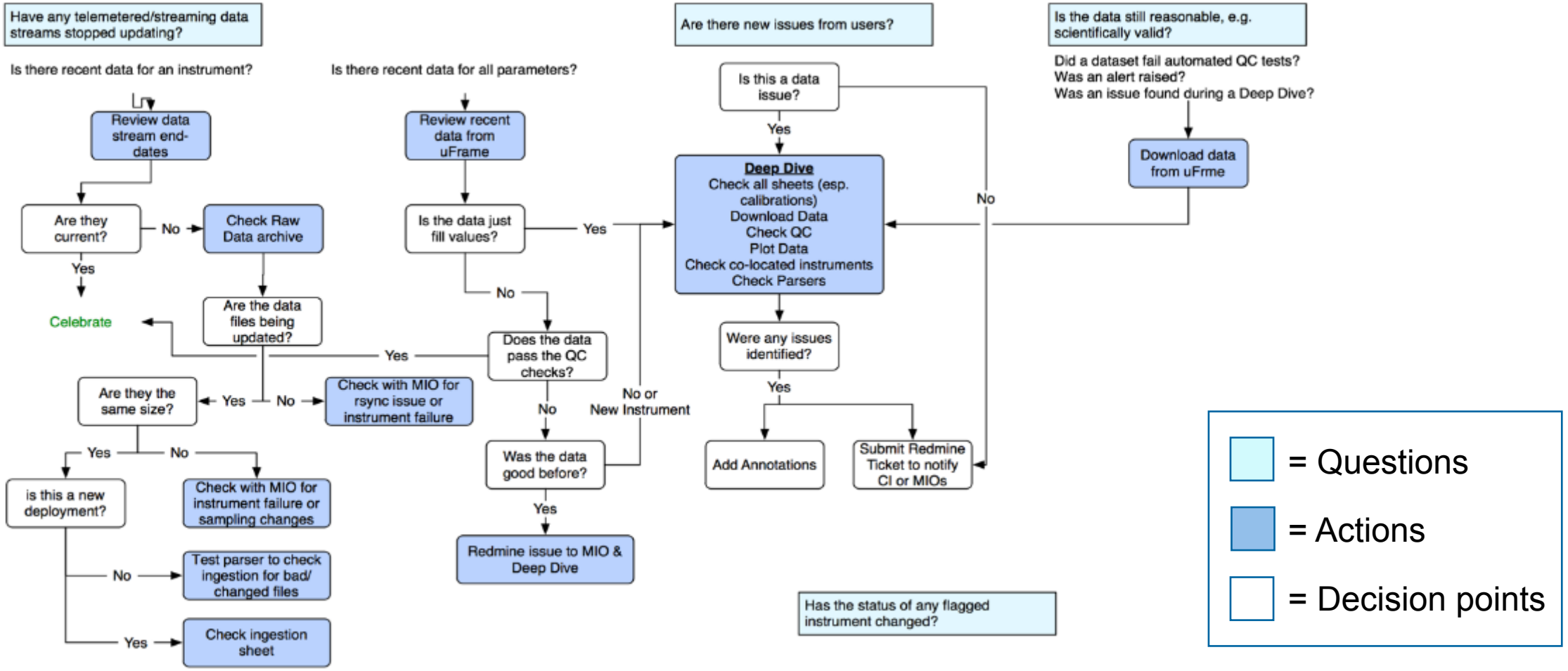


Data Evaluation Daily Activities



- 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

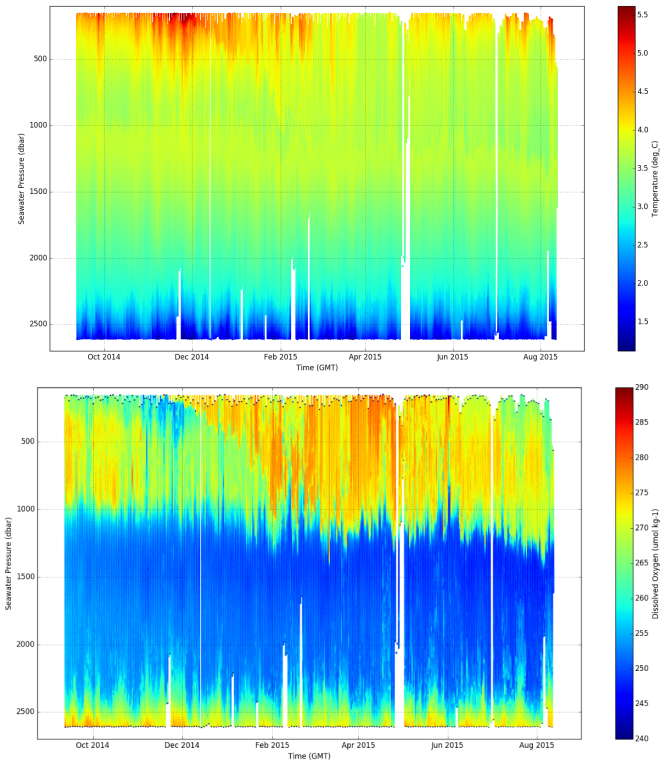


Periodic Data Team Activities

- 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

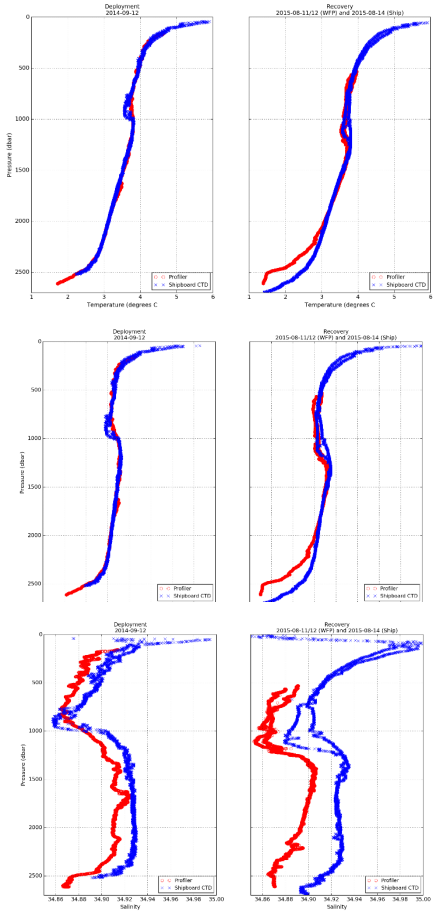
Data assurance/Data quality: *Pre and Post comparisons*

How well did the sensors fare over the time series?

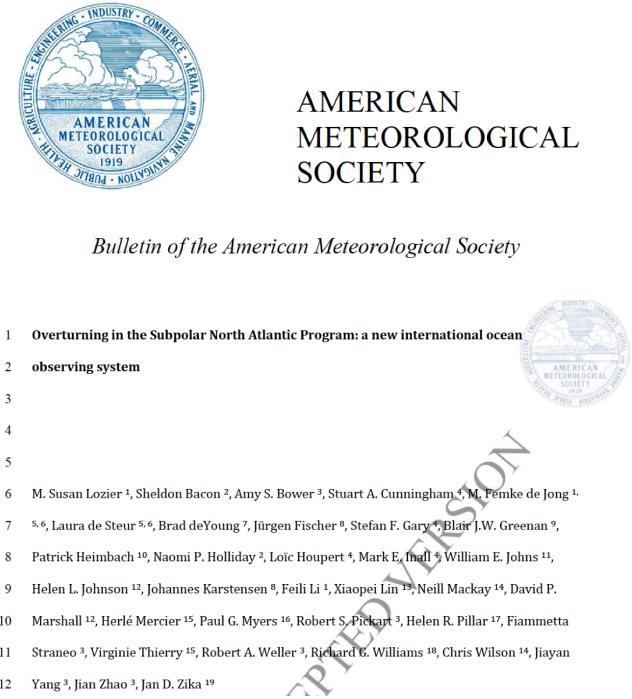


Profiler
Ship

Deployment *Recovery*



User applied post deployment QC



AMERICAN METEOROLOGICAL SOCIETY

Bulletin of the American Meteorological Society

Overturning in the Subpolar North Atlantic Program: a new international ocean observing system

M. Susan Lozier [1], Sheldon Bacon [2], Amy S. Bower [3], Stuart A. Cunningham [4], M. Femke de Jong [1, 5, 6], Laura de Steur [5, 6], Brad deYoung [7], Jürgen Fischer [8], Stefan F. Gary [9], Blair J.W. Greenan [9], Patrick Heimbach [10], Naomi P. Holliday [2], Loïc Houpert [4], Mark E. Inall [4], William E. Johns [11], Helen L. Johnson [12], Johannes Karstensen [8], Feili Li [1], Xiaopei Lin [13], Neill Mackay [14], David P. Marshall [12], Herlé Mercier [15], Paul G. Myers [16], Robert S. Pickart [3], Helen R. Pillar [17], Fiammetta Straneo [3], Virginie Thierry [15], Robert A. Weller [3], Richard G. Williams [18], Chris Wilson [14], Jiayan Yang [3], Jian Zhao [3], Jan D. Zika [19]

Global Array: Irminger

Overview

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

Sept. 2016 Workshop Feedback: Short Term SW fixes

i. No file **aggregation** prior to delivery. **Fixed**

ii. Improved **bathymetry**. **Fixed**

iii. Data Team **annotations** from MIO information about HW issues - **Fixed:** Data team now able to enter annotations, view them in GUI (more work to go)

iv. Large data download **time outs**, **request lost**, email response confusing. **Fixed**

v. Depth, Lat/long and pressure for all deployed instruments. **Fixed**

vi. Status timeline with metadata; overview from first deployment to present. Operator can perform manual updates. **Fixed**

vii. Incorporate all the naming and labeling options (**vocabulary**) that the data team added to preload, and improve filtering in the GUI. **Improved, but ongoing**

viii. Missing data & instruments in catalog (eg. MASSP, ZPLSC). **Improved but ongoing:** routes users to raw data archive, etc.; some analytical data still incoming

Workshop Feedback: Long-Term SW fixes

ix. Plotting from **multiple instruments** on 1 plot. **Fix:** Re-enabled, but needs GUI fix

x. Simplify Data Catalog to start with empty "cart" so users are not overwhelmed with options. **Future fix.**

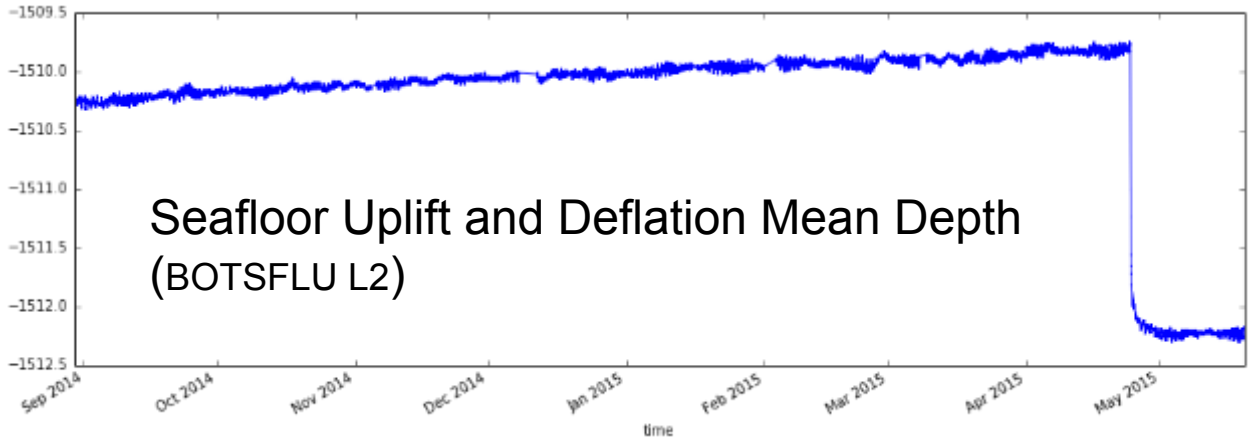
xi. Plotting clarifications: users should be informed if data cannot be plotted in 2D and best way to plot. **Future fix.**

xii. Improved links and access to raw data archive, documentation, and metadata info – **Fix in progress**

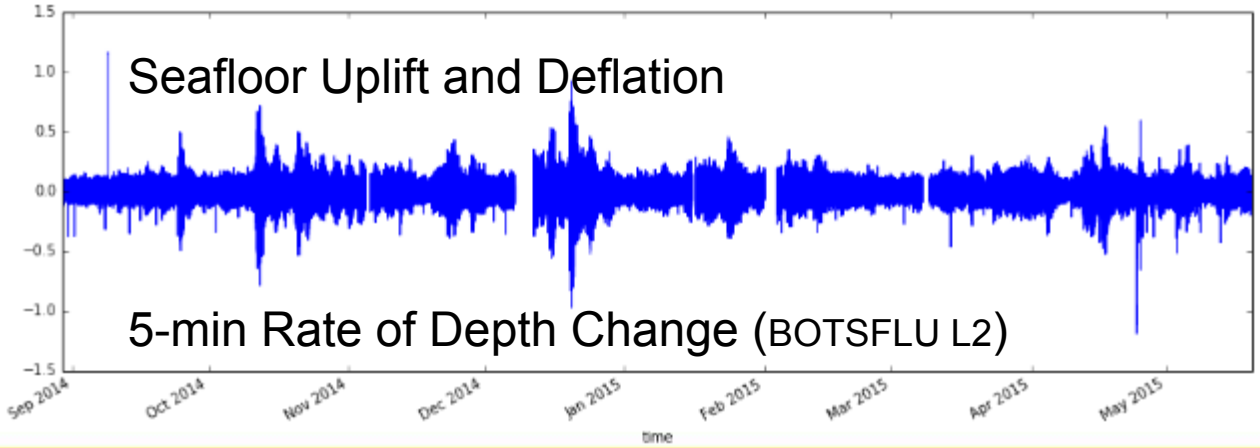
xiii. Improve Overall Data Quality. **Long-term fixes**: data team deep dives, secondary post-recovery calibration, external review by SMEs, Data Assembly Center (would require reallocation of funds)

Data Availability and Completeness

- Some derived data products are still being added to the system or require additional processing
- HPIES, ZPLSC/G, ADCP, MASSP, BOTPT, HYDBB, VADCP, FDCHP, CTDBP/FLORT, CAMDS, CAMHD, OSMOI, PPSDN, RASFL

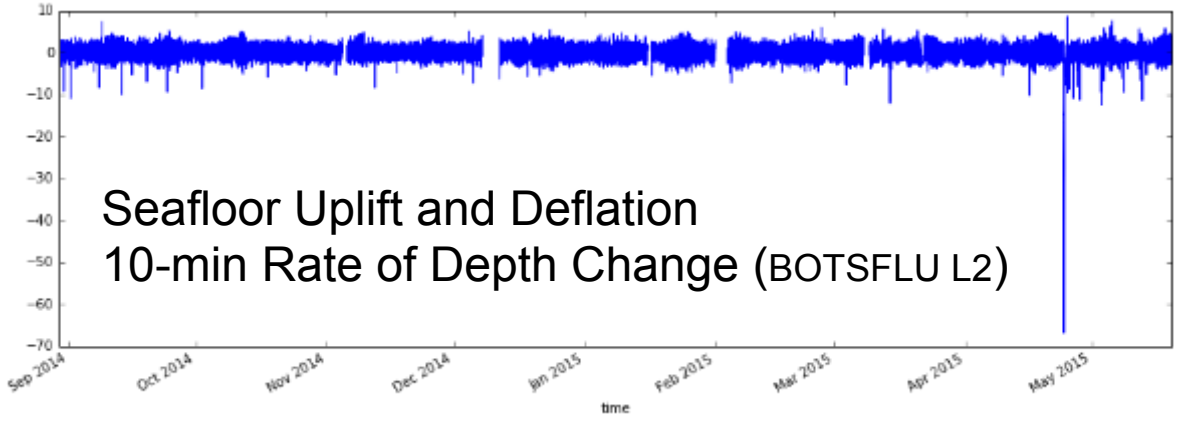


Seafloor Uplift and Deflation Mean Depth (BOTSFLU L2)



Seafloor Uplift and Deflation

5-min Rate of Depth Change (BOTSFLU L2)



Seafloor Uplift and Deflation
10-min Rate of Depth Change (BOTSFLU L2)

QC Challenges & Solutions

- Local range values need statistical analysis of environmental data for each platform
  - ***Need to work with SOC 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 SOC***
- 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***

Options for Data Review Acceleration

Option	Positive	Negative
MIO Operations Log at Rutgers	• Centralized log reduces time spent investigating issues • All issues entered consistently	• Takes time to maintain • Some development time
Speed up ingestion	• Fewer gaps to investigate	• Currently requires FTE • Automated process not yet delivered
Data Team works only on RIC	• More data reviewed faster	• No new data in system • No bug investigation • No QA testing
Limit reviewed time period or stream type	• Data reviewed slightly faster, at high level	• Review enhanced by looking at multiple deployments and trends • Slows down future reviews
Limit thoroughness of reviews	• Data reviewed faster, at high level	• Unclear why gaps exist • Quality issues not fully annotated • Slows down future reviews • Limits crowdsourcing options
Crowdsourcing (enlist volunteer SMEs)	• Removes subset of datasets from review queue • Assistance with complex data that requires expertise	• Focus on specific interest, not whole of OOI • Steep learning curve for advanced use of system (and knowledge of known issues) • Pathway to triage and incorporate feedback
Add employees or Data Assembly Center (DAC)	• Data reviewed faster, in depth • Support for expert analysis	• Requires additional funding • Setup and maintenance time

Adding capability to OOINet experience

Ocean Observatories » Web Interface

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

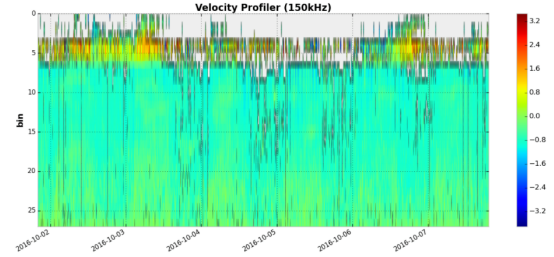
Enhancement #10913

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

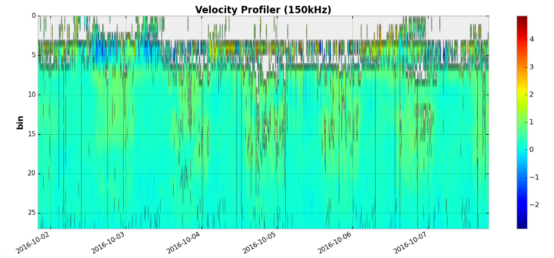
Added by Daniel Maher 4 months ago. Updated 16 days ago.

Status:	Ready for Work	**Start date:**	
Priority:	Urgent	**Due date:**	
Assignee:	Daniel Maher	**% Done:**	20%
Category:	-	**Spent time:**	-
Target version:	-		
Target Release:		**Array Affected:**	
Category 1:		**Instrument Affected:**	
Severity:		**CI Software Affected:**	
Issue Closed:		**Work Breakdown Structure (WBS):**	

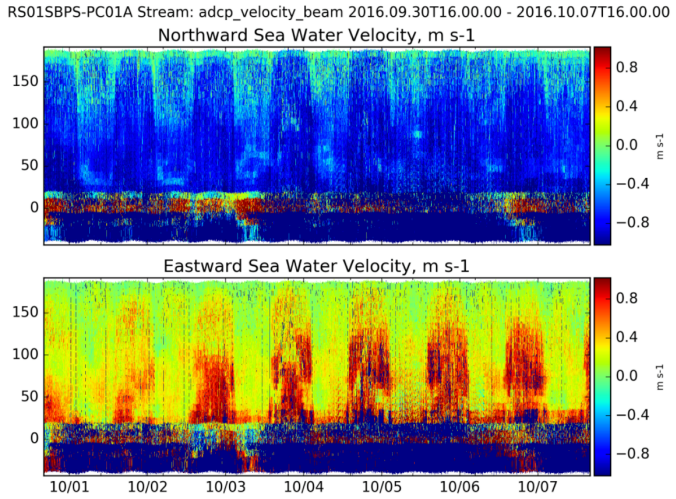
Northward UI Plot



Eastward UI Plot

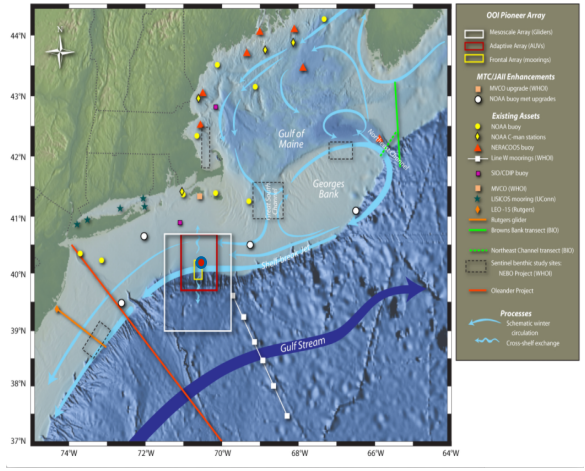


Python Plots

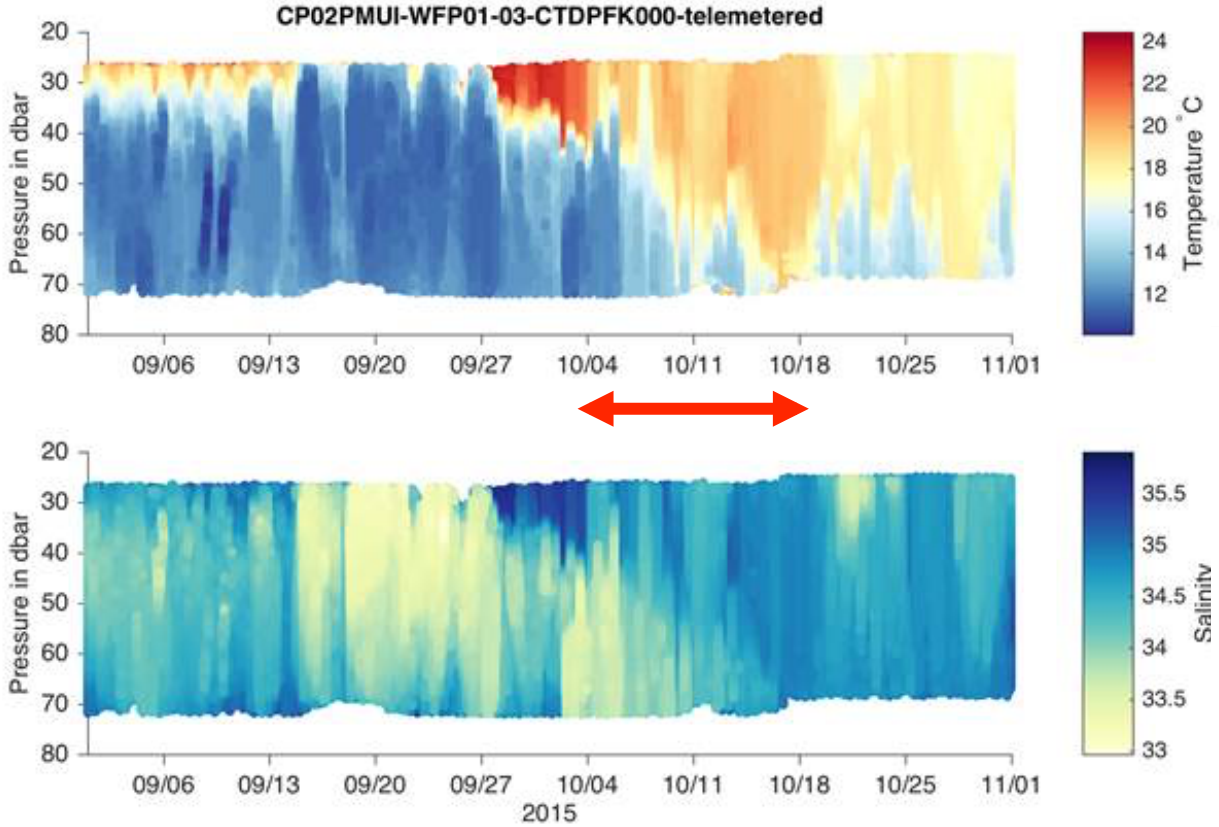


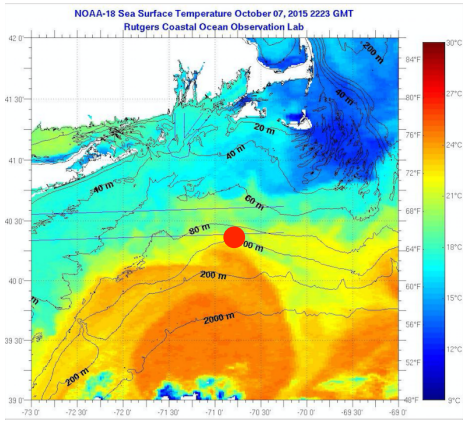
Science Evaluation: Are ocean features encountered real? Outside local range

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

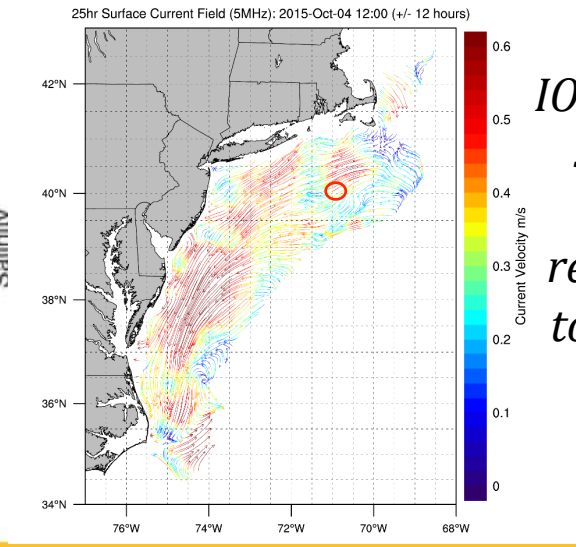


Is this real?



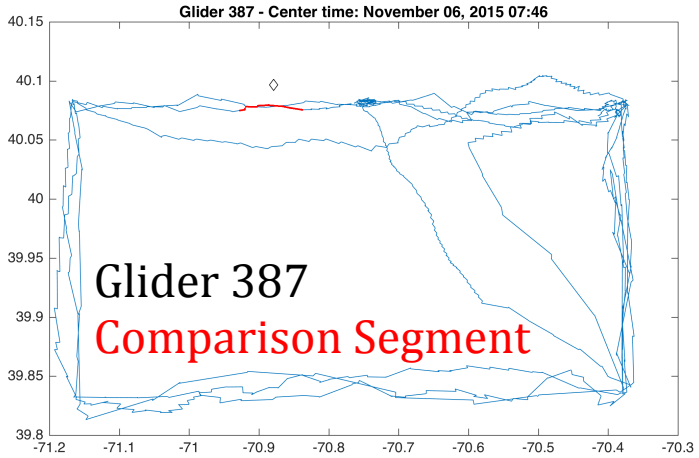


IOOS data shows presence of warm core ring

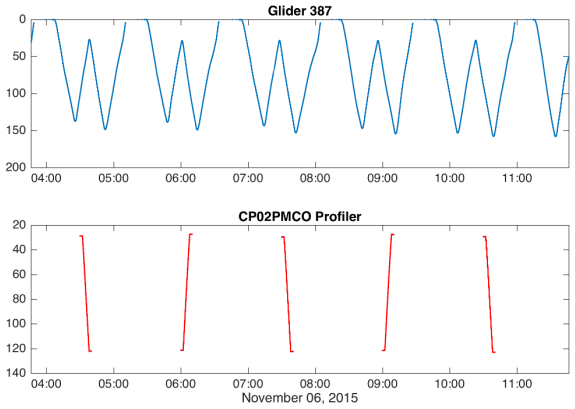


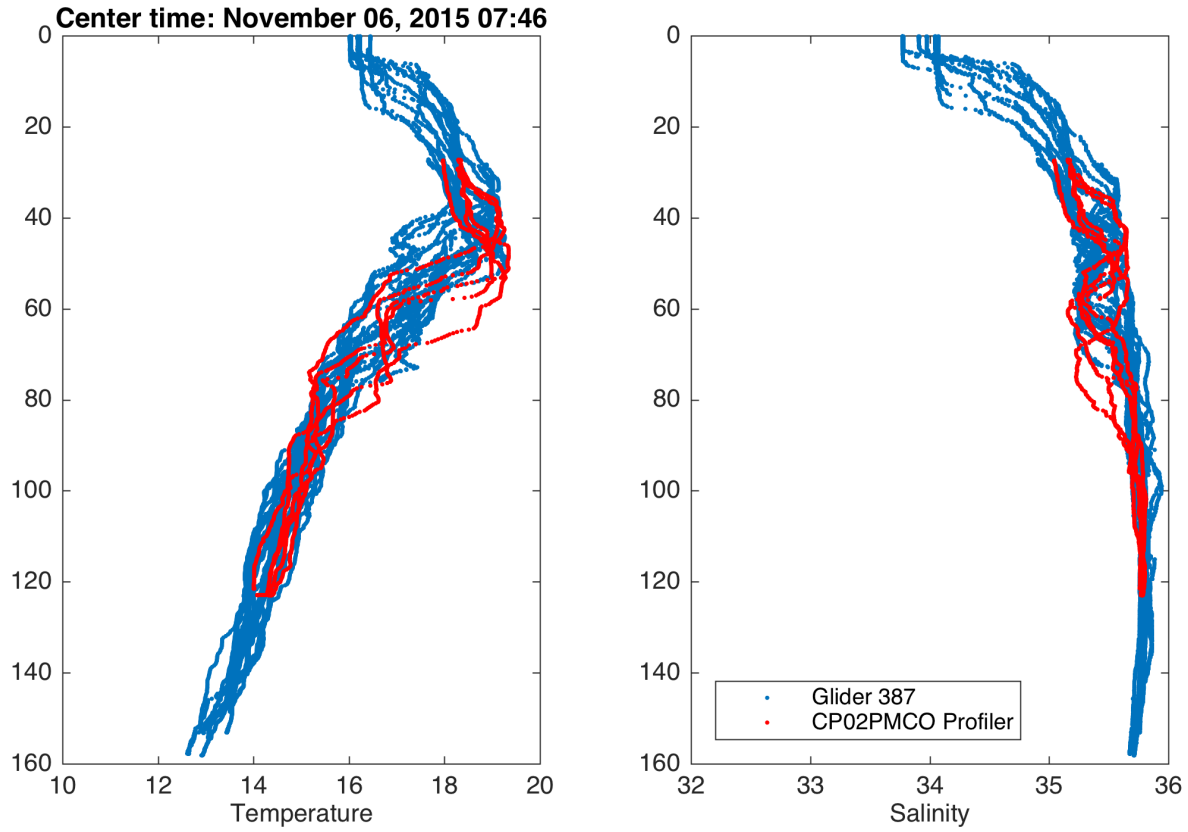
IOOS data shows ocean response to storm

Vicarious Calibration - Comparisons enabled by ERDDAP



Concurrent profiles from Glider 387 and Coastal Pioneer Profiler Mooring





Co-located and concurrent Temperature and Salinity Profiles

Blue – Glider 387 Red – Coastal Pioneer Profiler Mooring

Outreach and Community efforts

- Community tools section of website
- Hackathon to develop new visualization and processing tools
- Education meetings at Rutgers to get data into the classroom
- Data team presence at meetings and workshops (MTS, Ocean Sciences, Irminger)
- Engagement with SMEs

Overview

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

Conclusions

1. A large amount of high quality data has been and is being collected, with high science value
2. Data review is finally our primary focus, given maturation of the system
3. Data team accelerating RIC review via development of specialized tools
4. Short-term, medium-term, and long-term goals for improving data quality and delivery
5. OOI is providing a curated, consistent data system that is delivering data and metadata to the community

Questions?

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

Mike Vardaro, Data Manager, OOI CI Data Team
vardaro@marine.rutgers.edu

Mike Crowley, Program Manager, OOI CI Data Team
Crowley@marine.rutgers.edu

Acknowledgements: NSF, COL, Rutgers University, University of Washington, WHOI, Oregon State University, RPS-ASA, Raytheon, UCSD/SIO