Overview

1. Team Structure & Responsibilities
2. Data Flow & Products
3. Data Review
4. Communications
5. Improvements
6. Conclusions
First in Class Reviews: Jan-Aug 2016

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

<table>
<thead>
<tr>
<th>WBS</th>
<th>Task Name</th>
<th>% complete</th>
<th>Duration</th>
<th>Start</th>
<th>Finish</th>
</tr>
</thead>
<tbody>
<tr>
<td>290</td>
<td>1.5.1</td>
<td>Data Ingestion</td>
<td>62%</td>
<td>150 days</td>
<td>Wed 2/26/16</td>
</tr>
<tr>
<td>300</td>
<td>1.5.1.1</td>
<td>First in Class for Cassandra Team</td>
<td>88%</td>
<td>0 days</td>
<td>Mon 2/29/16</td>
</tr>
<tr>
<td>301</td>
<td>1.5.1.1.1</td>
<td>Pioneer Coastal Glider, CPO5MOAS-GL38B</td>
<td>100%</td>
<td>0 days</td>
<td>Mon 2/29/16</td>
</tr>
<tr>
<td>308</td>
<td>1.5.1.1.2</td>
<td>Pioneer Central Inshore Profiler Mooring, CPO3PACI</td>
<td>100%</td>
<td>0 days</td>
<td>Thu 1/28/16</td>
</tr>
<tr>
<td>315</td>
<td>1.5.1.1.3</td>
<td>Endurance OR Offshore Surface Moorings - CE095SM</td>
<td>100%</td>
<td>0 days</td>
<td>Mon 2/29/16</td>
</tr>
<tr>
<td>322</td>
<td>1.5.1.1.4</td>
<td>Called Slope Base Low Power Joos - RSO15LB-LTJ</td>
<td>84%</td>
<td>0 days</td>
<td>Thu 1/28/16</td>
</tr>
<tr>
<td>329</td>
<td>1.5.1.1.5</td>
<td>Called Slope Base Shallow Profiler Moorings - RSO11BS</td>
<td>100%</td>
<td>0 days</td>
<td>Thu 1/28/16</td>
</tr>
<tr>
<td>330</td>
<td>1.5.1.1.6</td>
<td>Called Slope Base Deep Profiler Moorings - RSO15SBPD</td>
<td>100%</td>
<td>0 days</td>
<td>Fri 3/11/16</td>
</tr>
<tr>
<td>343</td>
<td>1.5.1.1.7</td>
<td>Inning Global Ocean Glider, GD05SUMO</td>
<td>100%</td>
<td>0 days</td>
<td>Tue 2/26/16</td>
</tr>
<tr>
<td>350</td>
<td>1.5.1.1.8</td>
<td>Inning Sea Apex Profiler Moorings (GI02HFM)</td>
<td>100%</td>
<td>0 days</td>
<td>Tue 2/26/16</td>
</tr>
<tr>
<td>357</td>
<td>1.5.1.1.9</td>
<td>Inning Sea Flankers Subsurface Moorings A (GI05FLA)</td>
<td>69%</td>
<td>0 days</td>
<td>Wed 2/26/16</td>
</tr>
<tr>
<td>364</td>
<td>1.5.1.1.10</td>
<td>Inning Global Ocean Glider (G05MOAS-GL)</td>
<td>100%</td>
<td>0 days</td>
<td>Wed 2/26/16</td>
</tr>
<tr>
<td>371</td>
<td>1.5.1.1.11</td>
<td>Inning Global Profiling Gliders (G05MOAS-PG)</td>
<td>100%</td>
<td>0 days</td>
<td>Thu 2/26/16</td>
</tr>
<tr>
<td>378</td>
<td>1.5.1.1.12</td>
<td>Coastal Endurance OR Inshore Surface Piercing Profiler Moorings (CE011SSP)</td>
<td>100%</td>
<td>0 days</td>
<td>Fri 3/11/16</td>
</tr>
<tr>
<td>385</td>
<td>1.5.1.1.13</td>
<td>Coastal Endurance OR offshore BCP - CE040SBP</td>
<td>31%</td>
<td>0 days</td>
<td>Wed 3/26/16</td>
</tr>
<tr>
<td>392</td>
<td>1.5.1.1.14</td>
<td>Cabled Seafloor Instruments</td>
<td>0%</td>
<td>0 days</td>
<td>Mon 4/4/16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WBS</th>
<th>Task Name</th>
<th>% complete</th>
<th>Duration</th>
<th>Start</th>
<th>Finish</th>
</tr>
</thead>
<tbody>
<tr>
<td>418</td>
<td>1.5.2</td>
<td>Data Verification &amp; Validation</td>
<td>28%</td>
<td>255 days</td>
<td>Wed 2/10/16</td>
</tr>
<tr>
<td>419</td>
<td>1.5.2.1</td>
<td>Data Ingestion</td>
<td>99%</td>
<td>123 days</td>
<td>Tue 3/1/16</td>
</tr>
<tr>
<td>420</td>
<td>1.5.2.1.1</td>
<td>Pioneer Coastal Glider, CPO5MOAS-GL38B</td>
<td>95%</td>
<td>34 days</td>
<td>Tue 3/1/16</td>
</tr>
<tr>
<td>421</td>
<td>1.5.2.1.2</td>
<td>Endurance OR Offshore Surface Moorings - CE095SM</td>
<td>70%</td>
<td>30 days</td>
<td>Fri 3/4/16</td>
</tr>
<tr>
<td>422</td>
<td>1.5.2.1.3</td>
<td>Pioneer Upstream Inshore Profiler Moorings, CPO5PAP</td>
<td>47%</td>
<td>31.8 days</td>
<td>Fri 3/18/16</td>
</tr>
<tr>
<td>423</td>
<td>1.5.2.1.4</td>
<td>Called Slope Base Deep Profiler Moorings - RSO15SBP</td>
<td>0%</td>
<td>20 days</td>
<td>Thu 6/7/16</td>
</tr>
<tr>
<td>424</td>
<td>1.5.2.1.5</td>
<td>Called Slope Base Low Power Joos - RSO15LB-LTJ</td>
<td>0%</td>
<td>27 days</td>
<td>Fri 4/19/16</td>
</tr>
<tr>
<td>425</td>
<td>1.5.2.1.6</td>
<td>Called Slope Base Shallow Profiler Moorings - RSO15SB</td>
<td>75%</td>
<td>33 days</td>
<td>Thu 3/15/16</td>
</tr>
<tr>
<td>426</td>
<td>1.5.2.1.7</td>
<td>Inning Sea Apex Profiler Moorings (GI02HFM)</td>
<td>100%</td>
<td>20 days</td>
<td>Fri 3/18/16</td>
</tr>
<tr>
<td>427</td>
<td>1.5.2.1.8</td>
<td>Inning Global Open Ocean Glider (G05MOAS-GL)</td>
<td>0%</td>
<td>3 days</td>
<td>Fri 4/22/16</td>
</tr>
<tr>
<td>428</td>
<td>1.5.2.1.9</td>
<td>Inning Sea Flankers Subsurface Moorings A (GI05FLA)</td>
<td>0%</td>
<td>18 days</td>
<td>Wed 4/27/16</td>
</tr>
<tr>
<td>429</td>
<td>1.5.2.1.10</td>
<td>Inning Global Profiling Gliders (G05MOAS-PG)</td>
<td>0%</td>
<td>5 days</td>
<td>Mon 5/23/16</td>
</tr>
<tr>
<td>430</td>
<td>1.5.2.1.11</td>
<td>Inning (Or other global) Sea Apex Surface Moorings</td>
<td>40%</td>
<td>74 days</td>
<td>Mon 3/14/16</td>
</tr>
<tr>
<td>431</td>
<td>1.5.2.1.12</td>
<td>Coastal Endurance OR Inshore Surface Piercing Profiler Moorings (CE011SSP)</td>
<td>0%</td>
<td>8 days</td>
<td>Fri 6/24/16</td>
</tr>
<tr>
<td>432</td>
<td>1.5.2.1.13</td>
<td>Coastal Endurance OR offshore BCP - CE040SBP</td>
<td>0%</td>
<td>22 days</td>
<td>Thu 7/28/16</td>
</tr>
<tr>
<td>433</td>
<td>1.5.2.1.14</td>
<td>Called Axial Seamount Central Caldera Med Power J</td>
<td>0%</td>
<td>10 days</td>
<td>Fri 7/8/16</td>
</tr>
<tr>
<td>434</td>
<td>1.5.2.1.15</td>
<td>Cabled Seafloor Instruments</td>
<td>0%</td>
<td>20 days</td>
<td>Fri 7/22/16</td>
</tr>
<tr>
<td>435</td>
<td>1.5.2.1.16</td>
<td>ALUVs</td>
<td>0%</td>
<td>6 days</td>
<td>Tue 8/9/16</td>
</tr>
<tr>
<td>436</td>
<td>1.5.2.2</td>
<td>AGU Oceans Data Prep (THREDDS &amp; GUI) - Reasonability</td>
<td>100%</td>
<td>30 days</td>
<td>Wed 2/10/16</td>
</tr>
</tbody>
</table>
Data delivered:  T/S  R
Cabled       100  27*
Endurance   100  100

*Small subset of instruments are recovered only, need to be loaded
## QC Database: Higher Resolution Statistics

<table>
<thead>
<tr>
<th>Data Availability</th>
<th>T/S</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cabled</td>
<td>67</td>
<td>21</td>
</tr>
<tr>
<td>Endurance</td>
<td>64</td>
<td>53</td>
</tr>
<tr>
<td>Globals</td>
<td>47</td>
<td>56</td>
</tr>
<tr>
<td>Pioneer</td>
<td>77</td>
<td>50</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>57% T/S, 50% R</strong></td>
<td></td>
</tr>
</tbody>
</table>

*Some products are not expected due to platform loss, some recent deployments need ingestion

### Recovered

<table>
<thead>
<tr>
<th>Recovered Percentages by Array</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE</td>
</tr>
<tr>
<td>CA</td>
</tr>
<tr>
<td>GI</td>
</tr>
<tr>
<td>GP</td>
</tr>
<tr>
<td>RS</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
</tr>
</tbody>
</table>

### Telemetered/Streamed

<table>
<thead>
<tr>
<th>Telemetered/Streamed Percentages by Array</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE</td>
</tr>
<tr>
<td>CA</td>
</tr>
<tr>
<td>GI</td>
</tr>
<tr>
<td>GP</td>
</tr>
<tr>
<td>RS</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
</tr>
</tbody>
</table>

*Some products are not expected due to platform loss, some recent deployments need ingestion

**Data availability**

- **Telemetered/Streamed:**
  - QC: 67/72
  - CA: 83/90
  - GI: 65/68
  - GP: 85/85
  - RS: 52/48

- **Recovered:**
  - QC: 68/72
  - CA: 68/71
  - GI: 68/71
  - GP: 85/85
  - RS: 52/48

### Additional Notes

- **Endurance:**
  - Cabled: 64
  - Recovered: 53

- **Globals:**
  - Cabled: 47
  - Recovered: 56

- **Pioneer:**
  - Cabled: 77
  - Recovered: 50

### Total:

- **Cabled:** 67/21
- **Recovered:** 64/53

**Total:** 57% T/S, 50% R

**OOIFB Meeting:** May 2017
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

<table>
<thead>
<tr>
<th>Annotation ID</th>
<th>Annotation</th>
<th>Reference Designator</th>
<th>Stream Name</th>
<th>Start Date</th>
<th>End Date</th>
<th>Exclude Data?</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>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 7/12/18 onwards.</td>
<td>RS03AXBS-LJ03A-12-CTDPFB301</td>
<td>streamed_ctdpf-octode-sample</td>
<td>Tue, 12 Jul 2016 00:00:00 GMT</td>
<td>Thu, 01 Dec 2016 23:41:00 GMT</td>
<td>false</td>
</tr>
</tbody>
</table>
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

M2M Request → Plot Data File → Visual Assessment
- Data Availability
- Data Quality
→ Investigate Gaps and QC failures

Automated tests

Instrument Level
- Deployment time range

Stream Level
- Start and End Date
- Lat/Lon
- Duplicate Timestamps
- Parameter Availability

Parameter Level
- Global range
- Stuck Value
- Spike Test
- Fill values
- NaNs
- Gap Check

Summary Report

Annotate in GitHub
- Subsite – Operational Status
- Node – Operational Status
- Instrument – Operational Status
- Stream - Availability
- Parameter – Quality & Availability

Push to System
Automated Scripting Tools

- **parse_spring_files**: Used to parse data into an easy to use lookup table that routes each spring file to its proper uframe_route and driver

- **check_ooi_nc**: This toolbox is used by the OOI Data Review Team at Rutgers University in order to check netCDF files for accuracy

- **plot-nc-ooi**: Python script used to plot OOI netCDF datasets

- **ooi-parameters-dict**: A repository containing a dictionary of streams and corresponding science parameters

- **list_omc**: Used to check updates to Raw Data directory (data team only; password protected)
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)
<table>
<thead>
<tr>
<th>OOI Test</th>
<th>OOI Description</th>
<th>QARTOD Equivalent</th>
<th>QARTOD Recommendation (from manuals)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Range Test</td>
<td>Data are flagged unless they fall within valid world ocean ranges or instrument limits</td>
<td>Gross Range</td>
<td>Only considers manufacturer-defined sensor and calibration limits</td>
<td>Different tests, different names. Currently operational.</td>
</tr>
<tr>
<td>Spike Test</td>
<td>Deviation from mean compared to 2*N neighboring points</td>
<td>Spike</td>
<td>N=1, default threshold is based on the rate of change distribution from previous data sets</td>
<td>Roughly identical, same nomenclature. Currently operational.</td>
</tr>
<tr>
<td>Stuck Value Test</td>
<td>If 2 neighboring values differ by less than the resolution of the sensor for more than N repetitions, data are flagged</td>
<td>Stuck Sensor</td>
<td>Manual suggests 3 consecutive points for a stuck sensor suspect flag and 5 for a fail flag.</td>
<td>QARTOD manual suggestion may be too low for well-mixed portions of the water column. Under evaluation.</td>
</tr>
<tr>
<td>Local Range Test</td>
<td>Data are flagged unless they fall within locally valid site-specific or depth ranges. Interpolates thresholds between depth and season intervals</td>
<td>Local Range</td>
<td>Starts with constant limits for each depth/season interval</td>
<td>Roughly identical, same nomenclature. OOI Local ranges are still being established.</td>
</tr>
<tr>
<td>Gradient Test</td>
<td>If d(data)/d(t) &gt; 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.</td>
<td>Rate of Change</td>
<td>QARTOD recommends two neighboring points and does not incorporate TOLDAT or STARTDAT values.</td>
<td>Different tests, different names. Under Evaluation, not operational</td>
</tr>
<tr>
<td>Trend Test</td>
<td>Data flagged as having trend if the SD of the residuals to a polynomial curve &lt; original data, multiplied by a factor. Test for sensor drift.</td>
<td>N/A</td>
<td>No QARTOD equivalent</td>
<td>OOI only. Under Evaluation, not operational</td>
</tr>
</tbody>
</table>
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*
## Rest in Class Data Status Categories

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

Quality Timeline

Annotation Text

May 2017 OOIFB Meeting
QC Database Tool

- Used for reference & statistics
- Includes status information
- Includes testing/review capability
- Annotation
- http://ooi.visualocean.net
Post-Cruise/Post-Deployment Checklist

- Part of Rest-in-Class review: vital annotation information, as well as data delivery to users
- Used as a pre-cruise and post-cruise check on shipboard data and documentation
- Also useful to determine % completeness of data delivery
- Can enter notes or indicate whether completion is blocked by delivery of a dataset or document
Deliverables

• Data Availability Reports
  o (% completeness, streams/parameters reported, particles in the system)
• Data Quality Reports
• Redmine reporting
  o Issues found, investigations, and Help Desk open/closed
• Deep dive investigation reports
• Annotations (to users)
• Download statistics
• Forum statistics (TBD)
<table>
<thead>
<tr>
<th>Option</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIO Operations Log at Rutgers</td>
<td>• Centralized log reduces time spent investigating issues</td>
<td>• Takes time to maintain</td>
</tr>
<tr>
<td></td>
<td>• All issues entered consistently</td>
<td>• Some development time</td>
</tr>
<tr>
<td>Speed up ingestion</td>
<td>• Fewer gaps to investigate</td>
<td>• Currently requires FTE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Automated process not yet delivered</td>
</tr>
<tr>
<td>Data Team works only on RIC</td>
<td>• More data reviewed faster</td>
<td>• No new data in system</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No bug investigation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No QA testing</td>
</tr>
<tr>
<td>Limit reviewed time period or stream type</td>
<td>• Data reviewed slightly faster, at high level</td>
<td>• Review enhanced by looking at multiple deployments and trends</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Slows down future reviews</td>
</tr>
<tr>
<td>Limit thoroughness of reviews</td>
<td>• Data reviewed faster, at high level</td>
<td>• Unclear why gaps exist</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Quality issues not fully annotated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Slows down future reviews</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Limits crowdsourcing options</td>
</tr>
<tr>
<td>Crowdsourcing (enlist volunteer SMEs)</td>
<td>• Removes subset of datasets from review queue</td>
<td>• Focus on specific interest, not whole of OOI</td>
</tr>
<tr>
<td></td>
<td>• Assistance with complex data that requires expertise</td>
<td>• Steep learning curve for advanced use of system (and knowledge of known issues)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Pathway to triage and incorporate feedback</td>
</tr>
<tr>
<td>Add employees or Data Assembly Center (DAC)</td>
<td>• Data reviewed faster, in depth</td>
<td>• Requires additional funding</td>
</tr>
<tr>
<td></td>
<td>• Support for expert analysis</td>
<td>• Setup and maintenance time</td>
</tr>
</tbody>
</table>
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

- Have any heterogeneous/streaming data streams stopped updating?
- Is there recent data for an instrument?
- Review data stream end-dates
- Are they current?
- Yes
  - No
    - Check Raw Data archive
    - Are the data files being updated?
    - Yes
      - No
        - Check with MIO for instrument failure
        - Was the data good before?
          - Yes
            - Test parser to check ingestion for bad/changed files
            - Check Ingestion sheet
          - No
            - Redline Issue to MIO & Deep Dive
    - No
      - Check with MIO for instrument failure or sampling changes
- No
  - Yes
    - Is the data just filled values?
    - Yes
      - No
        - Does the data pass the QC checks?
          - Yes
            - Add Annotations
          - No
            - No or New Instrument
    - No
    - Review recent data from uFrame
      - Is the data just filled values?
        - Yes
          - No
            - Does the data pass the QC checks?
              - Yes
                - Submit Redline Ticket to notify CI or MIOs
              - No
                - No or New Instrument
            - No
              - Is there new issues from users?
                - Yes
                  - No
                    - Is the data still reasonable, e.g., scientifically valid?
                      - Yes
                        - No
                          - Did a dataset fail automated QC tests?
                            - Yes
                              - No
                                - Was an alert raised?
                                  - Yes
                                    - No
                                      - Was an issue found during a Deep Dive?
                                        - Yes
                                          - No
                                            - Download data from uFrame
                                          - No
                                            - Check Co-located Instruments
                                              - Check Failsafe
                                              - Check with MIO for QC issues or instrument failure
                                              - Check with MIO for new deployment
                                              - Check with MIO for instrument failure or sampling changes
                                              - Test parser to check ingestion for bad/changed files
                                              - Check Ingestion sheet
                                              - Is there recent data for all parameters?
                                                - Yes
                                                  - No
                                                    - Review recent data from uFrame
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*

**Global Array: Irminger**

How well did the sensors fare over the time series?

**Profiler**

**Deployment**

**Recovery**

User applied post-deployment QC

**May 2017 OOIFB Meeting**