Cyberinfrastructure 1.0 Status Update
OOI Facility Board Mtg, May 7-8, 2018

Orest E. Kawka

OOI Data Integrator (Jan 1 - Sept 30th, 2018)
Consortium for Ocean Leadership
&
Project Scientist
Cabled Array, University of Washington
Recurring CI-Related Activities

• **Weekly CI Coordination Call (Wed)**
  - Redmine Ticket Review: New Tickets, Pop-up Task Identification, Reprioritization, Sprint Checkup
  - Participants: Rutgers CI (Kish, Rodero), Rutgers Data Team (Crowley, Vardaro), Data Integrator (Kawka)

• ~ **Weekly – Biweekly Data Team Management + DI Call**
  - Participants: Rutgers Data Team (Crowley, Vardaro), OOI Data Integrator (Kawka)
  - Discuss Data Team Activities, Data Quality Issues and Path to Resolution

• ~ **Biweekly MIO POCs + DI Call (Started - April 18)**
  - Participants: Endurance (Risien), Coastal Global (Petillo), OOI Data Integrator (Kawka)
  - Discuss MIO-associated data flow issues, configuration data and other metadata availability entry, etc.

• **Weekly COL CI+Transition Update**
  - Participants: OOI Director (Rutherford), OOI Transition Manger (Kearney), OOI Data Integrator (Kawka)
  - Discuss MIO-associated data flow issues, configuration data and other metadata availability entry, etc. etc.
Recurring CI-Related Activities (cont)

- **Weekly CI Transition (TDP) Working Group**
  - Participants: CI (Rodero), Data Team (Crowley, Vardaro), COL (Kawka, Kearney), SMEs (Fram, Feldman, Petillo)
  - Status of CI Transition and TDP, Identify Required+Missing Documentation, Review TDP
  - Goal: Provide sufficient level of documentation to enable OOI 2.0 to Manage CI and Deliver Quality Data to Community
Current Status of Data Delivery

- Data Availability Plots are now on OOINet at instrument level
- Daily and Monthly Stats for Data Coverage are available at Data Team–supported site: ooi.visualocean.net
- Daily Alerts via email to Marine Operators re: Data Stream Issues (Knuth code)
- OOI Gliders - 103 are now available in IOOS Glider DAC
- Data Team is continuing Data Evaluation/Validation
  - Total of 1196 Data Annotations available on production server, available in UI plotting and downloadable JSON files
  - All annotations for existing issues caught up
  - MIO-Based upload of cruise data critical to evaluation is continuing
- Data Team continues conducting ingestion process of all Uncabled Data
  - Primary party for identifying issues with ingestion process and correcting via Redmine ticket issuance, testing, and resolution
  - Still need full instructions and functionality for playback of data
- Data Team continues to identify significant issues with MIO-supplied configuration data (e.g. calibrations) and other metadata
  - Significant impact on validity of data products currently being released to the community.
Current Status of Data Delivery (Cont)

• Global, Local, Spike, Stuck Thresholds –
  • Unit inconsistencies in Preload for 2500+ instrument parameters identified and being updated
  • Sprint 13 fix will address issues with QC Database and Global Range QC mismatch
  • Local Updates, Stuck and Spike pushed to OOI 2.0 due to Rutgers Data Team attrition and additional time needed to identify gaps and organize OOI cruise data.
  • Time Remaining in 1.0 - Will be used to run high-level QC on all data within system

• ERDDAP Server
  • 2256 Uncabled (Telemetered) Data Streams Available - 439 of those are current deployments
  • Moving ahead with update to ERDDAP Architecture - should improve function for high-density Cabled Array Data and reduce storage needs by ~50%
  • ERDDAP Development Cycle – targeted as a 1-week cycle rather than 2-week cycle
  • **Goal by Sept. 30th:** All OOI Cabled and Uncabled Recovered Datasets will be added to repository
  • Do we need to determine how to propagate Data Annotations to ERDDAP? Also no provenance in ERDDAP
**CI Software + Redmine Ticket Status**

**Sprint 13 Demo conducted Thursday, April 26th**
- Redmine Ticket Review: New Tickets, Pop-up Task Identification, Reprioritization, Sprint Checkup
- Participants: Rutgers CI (Kish, Rodero), Rutgers Data Team (Crowley, Vardaro), Data Integrator (Kawka)

**Sprint 13 Deployment to Production Scheduled for Tuesday, May 8th**
- Redmine Tickets Addressed: 27 Tickets
  - Rutgers Data Team will continue evaluating ticket resolution on production server before closing tickets

**Sprint 14 (Current)**
- Redmine Tickets in progress: 29
- Reprioritization due to Pop-Up Tasks during weekly CI Coordination Calls (CI admin+Data Team Management+Data Integrator)
- Rutgers Data Team - Primary group for critical testing of fixes/enhancements during development cycle and final integration/testing/demo

**Remaining Redmine Tickets ~600**
- 77 of those can probably be closed out
- Some of remaining tickets are subtasks, so actual number may be a bit lower
- Mix of support (help desk), bugs, enhancements spread across Uframe, GUI, instrument configuration data and other metadata etc. (details later)
CI Software Schedule (Prospective)

- **Sprint 12** (completed)
- **Sprint 13** (completed)
- **Sprint 14** (current)
- **Sprint 15**
- **Sprint 16**
- **Sprint 17**

**Legend:**
- Green: Completed
- Orange: Pending (current target)
- Blue: To be scheduled
- Light Blue: Deployment in production
- Black Diamond: Integration testing and Demo

Sprint 13 deployment!!
Redmine Ticket Backlog Breakdown (657 total)

**Sprint 13 (Deploy 5-8-18) – 27 RT**
- 4% Invalid Science Data
- 22% Missing Science Data
- 4% Missing Eng Data
- 30% uFrame Bugs
- 15% uFrame Enhancement*
- 22% UI Bugs
- 4% UI Enhancement

*Enhancement - a bit of a misnomer

**Sprint 14 (Current) – 29 RT**
- 3% Invalid Science Data
- 17% Missing Science Data
- 41% uFrame Bugs
- 17% uFrame Enhancement
- 10% UI Bugs
- 7% UI Enhancement
- 3% Metadata

**Remaining Backlog ~ 600 RT**
~13% Could be closed

Potential Increase w/ Data Issues as more Validation occurs!!
### Current Sprint 14 Tickets (29 total)

<table>
<thead>
<tr>
<th>#</th>
<th>Project</th>
<th>Tracker</th>
<th>Status</th>
<th>Priority</th>
<th>Subject</th>
<th>Affected Instrument</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>13246</td>
<td>OOINet Integration</td>
<td>Bug</td>
<td>In Progress</td>
<td>Normal</td>
<td>Annotation ID 01 for a BOTPT re-leveling event does not appear in UI</td>
<td>All</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13231</td>
<td>Web interface</td>
<td>Bug</td>
<td>New</td>
<td>Normal</td>
<td>Panor stuck in loop</td>
<td>All</td>
<td>Sprint 14, U bug</td>
</tr>
<tr>
<td>13221</td>
<td>OOINet Integration</td>
<td>Bug</td>
<td>Feedback</td>
<td>Normal</td>
<td>Stream Engine resolves parameters not requested</td>
<td>All</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13191</td>
<td>uFrame</td>
<td>Bug</td>
<td>New</td>
<td>Normal</td>
<td>Ensure that instruments are not using old calibration values</td>
<td>All</td>
<td>Sprint 14, U bug</td>
</tr>
<tr>
<td>13181</td>
<td>OOINet Integration</td>
<td>Bug</td>
<td>New</td>
<td>High</td>
<td>Stream engine unable to interpretate secondary dataset</td>
<td>Multiple</td>
<td>WACCP Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13166</td>
<td>Web interface</td>
<td>Bug</td>
<td>New</td>
<td>Normal</td>
<td>Unable to find annotations by partial reference designator using m2r</td>
<td>Multiple</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13147</td>
<td>OOINet Integration</td>
<td>Bug</td>
<td>New</td>
<td>High</td>
<td>Unable to playback data from Port Agent Log files</td>
<td>Multiple</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13145</td>
<td>OOINet Integration</td>
<td>Support</td>
<td>Feedback</td>
<td>High</td>
<td>Unable to map reference designators to sensors or sensors via an M2M protocol</td>
<td>Cabled</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13113</td>
<td>OOINet Integration</td>
<td>Enhancement</td>
<td>In Progress</td>
<td>Normal</td>
<td>Provide Missing Data to Algorithms</td>
<td>All</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13111</td>
<td>Instrument Agent Drive</td>
<td>Enhancement</td>
<td>In Progress</td>
<td>High</td>
<td>Improve Stream Engine Interception</td>
<td>All</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13101</td>
<td>Instrument Agent Drive</td>
<td>Bug</td>
<td>In Progress</td>
<td>High</td>
<td>UFrame bug</td>
<td>All</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13100</td>
<td>uFrame</td>
<td>Bug</td>
<td>New</td>
<td>Normal</td>
<td>UFrame supports data in 18 MHz and 802 KHz chunks, instead of 90 MHz or by deployment</td>
<td>All</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13082</td>
<td>Instrument Agent Drive</td>
<td>Bug</td>
<td>Feedback</td>
<td>High</td>
<td>UFrame calibrate outside stream time bounds</td>
<td>Cabled</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13060</td>
<td>uFrame</td>
<td>Enhancement</td>
<td>Feedback</td>
<td>High</td>
<td>UFrame calibrate outside stream time bounds</td>
<td>Cabled</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>13040</td>
<td>Instrument Agent Drive</td>
<td>Enhancement</td>
<td>Feedback</td>
<td>High</td>
<td>UFrame calibrate outside stream time bounds</td>
<td>Cabled</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>12975</td>
<td>Web interface</td>
<td>Bug</td>
<td>New</td>
<td>Urgent</td>
<td>Data availability timeline in production showing data available for a time range for which no data exists</td>
<td>DO</td>
<td>Sprint 14, missing data</td>
</tr>
<tr>
<td>12840</td>
<td>Web interface</td>
<td>Bug</td>
<td>New</td>
<td>Progress</td>
<td>RUN Data availability timeline displaying &quot;Present&quot; data where valid deployment data are not available</td>
<td>All</td>
<td>Sprint 14, U bug</td>
</tr>
<tr>
<td>12770</td>
<td>Web interface</td>
<td>Bug</td>
<td>New</td>
<td>Normal</td>
<td>FTP-X fails on annotation POST when an &quot;if&quot; element is presented in the input data</td>
<td>Multiple</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>12570</td>
<td>OOINet Integration</td>
<td>Bug</td>
<td>Feedback</td>
<td>High</td>
<td>NetCDF download missing data points</td>
<td>Cabled</td>
<td>BOTPT Sprint 14, missing data</td>
</tr>
<tr>
<td>12540</td>
<td>OOINet Integration</td>
<td>Bug</td>
<td>Feedback</td>
<td>High</td>
<td>Asset Management Cased Noise Reference Designators are not being fully qualified</td>
<td>Cabled</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>12525</td>
<td>Web interface</td>
<td>Bug</td>
<td>Enhancement</td>
<td>Progress</td>
<td>Data catalog in U should be collapsed to display a single entry per instrument</td>
<td>Cabled</td>
<td>Sprint 14, U enhancement</td>
</tr>
<tr>
<td>12510</td>
<td>Web interface</td>
<td>Bug</td>
<td>Progress</td>
<td>New</td>
<td>UFrame Annotation - Cal annotations by peripheral reference designator (platform and/or node) fail</td>
<td>Cabled</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>12316</td>
<td>Web interface</td>
<td>Enhancement</td>
<td>Feedback</td>
<td>High</td>
<td>Add cable lay position and other seafloor cable infrastructure to landing page and platform maps on U</td>
<td>Cabled</td>
<td>BOTPT Sprint 14, missing data</td>
</tr>
<tr>
<td>12226</td>
<td>OOINet Integration</td>
<td>Bug</td>
<td>Enhancement</td>
<td>Progress</td>
<td>BOTPT Transducer temperature parameter value being truncated by driver/processor</td>
<td>Cabled</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>12185</td>
<td>OOINet Integration</td>
<td>Bug</td>
<td>Enhancement</td>
<td>Progress</td>
<td>BOTPT Transducer temperature parameter value being truncated by driver/processor</td>
<td>Cabled</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>9990</td>
<td>Instrument Agent Drive</td>
<td>Bug</td>
<td>Feedback</td>
<td>High</td>
<td>VACCP - missing instrument parameters and failed creation of the uplink winery</td>
<td>Cabled</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>9880</td>
<td>Instrument Agent Drive</td>
<td>Bug</td>
<td>In Progress</td>
<td>Normal</td>
<td>UFrame bug</td>
<td>Cabled</td>
<td>Sprint 14, uFrame bug</td>
</tr>
<tr>
<td>8662</td>
<td>Instrument Agent Drive</td>
<td>Bug</td>
<td>In Progress</td>
<td>Normal</td>
<td>Add CTDer and FLOT to OES16SSM &amp; CE616SSM per ECR 1300-03530</td>
<td>Cabled</td>
<td>Sprint 14, uFrame bug</td>
</tr>
</tbody>
</table>
Other Current and Upcoming Activities

• **CI Requirements Refresh**
  • Contractor: Raytheon; scheduled to be completed by Aug 15th
  • Will provide CI System Design and Specification documentation

• **Formalized CI Change Control Board (CCB) in process of being formed**
  • Will be used, as necessary, for final approval of documentation for the OOI CI TDP
  • Engineering Change Requests (ECRs) for configuration and architecture changes, algorithms, etc

• **Management of Data Requests**
  • During UW Cabled Array Hackathon, significant delays in data downloads
  • Short-Term and Long-Term Solutions have been proposed (Short-Term Recommendations Paper Forthcoming)
  • Long-Term Solution: Development of a CI Request Manager in uFrame
Other Current and Upcoming Activities (cont)

• **UI Assessment**
  - Science Gateways Community Institute (NSF-funded): Purdue University staff
  - Limited Scope

• **OOI Early Career Workshops (4)**
  - Physics (May 21-25), Biology (June 18-22), Geology (July 16-20), Chemistry (July 30-Aug 3)
  - Requires significant resources from Rutgers Data Team to train 40-50 early career scientists and support workshops

• **CI Activity Prioritization and Resource Reallocation for Transition to OOI 2.0**
  - Ongoing discussions
Ongoing Issues and Risks

• Significant CI software fixes work remains (> 60% of remaining 500 Redmine Tickets)
  • Critical to fix “bad” and missing data

• Continuing issues with configuration data (instrument identifications, calibrations) and other metadata
  • Invalid data has been released to the community!!!
  • Currently no method exists to notify users of “bad” data downloaded
  • Data Team is primary resource for identifying these issues and notifying the MIOs

• Validation of existing data by Data Team is an ongoing process, labor intensive
  • Uses QC algorithm flags and multitude of tools developed by the Data Team
  • Comparison with cruise-based verification data is ongoing – has identified issues with ingestion, algorithms, and metadata

• Significant Risk - Data Team overload and attrition during remainder of OOI 1.0 (ends Sept. 30th)
  • Critical link in Redmine fix/development process: Primary group of testers in development cycle
  • Requested to provide support for CI TDP development
  • Significant support for four (4) upcoming Early Career Scientist Workshops
  • 25% of Data Analysts leaving early July; Additional attrition expected over next few months.