Regional Cabled Array (RCA)  
Current Data QA/QC Activities and Priorities

Orest E. Kawka

Research Scientist, OOI Regional Cabled Array  
School of Oceanography  
University of Washington  
Seattle, WA 98195

OOIFB and DDCI Joint Meeting, October 30, 2018
RCA: Current Data Processes

- **Instrument Data Ingestion**
  - Automatically parsed into CASSANDRA upon arrival at the servers on shore at full temporal resolution. No post-recovery ingestion needed.
    - Investigating issues with Deep Profiler data ingestion
  - Raw Data & Core Analytical Products are rsynced to OOI Raw Data Server from OMC repository at UW
  - Interruptions in ingestion process require back-filling of gaps using purge & playback from OOI Raw Data Server or Port Agent Logs

- **Instrument Port Agents and Driver/Parser Servers**
  - In OOI 2.0 RCA Team takes over responsibility
    - Six Servers moved from Pittock Bldg (Portland, OR) to 4545 Bldg, UW
    - Servers will eventually move to Shore Station in Pacific City, OR
    - Management hand off pending additional action by CI/Raytheon
    - To be developed: Interface Control Document (ICD)
RCA Data QA/QC Activities

• **Current Ongoing Priorities**
  - Continuing QA/QC work of 1.0 Rutgers Data Team
  - Working existing Redmine Tickets for ongoing data issues
  - Addressing new data issues as they are identified by internal (UW/APL engineers) and external (HelpDesk, Iris, etc.) monitoring
  - Downstream data QA/QC using field verification data and comparison of colocated instruments
  - Reviewing/updating existing and adding new data annotations
  - Data Algorithm updates as needed to resolve data issues
    - pCO$_2$ in water correction: Needs final testing and deployment
  - Exploring existing tools and scripts provided by 1.0 Rutgers Data Team
  - Asset Management Updates (*Critical Priority*)

• **OOI 2.0 Data QA/QC Plan & Procedures**
  - Updating and improving on 1.0 QA/QC procedures in coordination with OOI PMO and other MIOs
RCA Data QA/QC Activities

• **Asset Management Uploads**
  - Input into CI system of all metadata associated with OOI infrastructure (e.g. instruments) during refurbishment cycle
  - Continuing established process per MIO responsibilities prior to transition to OOI 2.0

• **New 2.0 Responsibility** – Final QC checks of all instrument associated metadata post-upload, including “Critical Metadata”
RCA Data QA/QC Activities

- “Critical Metadata”
  - Instrument Calibration Data, e.g. coefficients
  - Instrument Assignments – Deployment Sheets: What is where?

- Current High Priority: Check of historical critical metadata prior to 2.0 (2013-2018)
Comprehensive End-to-End Instrument and Data QA/QC

Quick View

Instrument Information

- Instrument Service and Testing Process
- Critical Metadata QA/QC Process
- Asset Management
- Raw Data and Data Products
- Data Product QA/QC Process

Instrument Life Cycle

- Instrument Integration Process
- Shipboard Deployment and Recovery
- Shore Initialization and Checkout

Legend:
- RCA Data Team
- RCA Engineering Team
- OOI CI
RCA Instrument and Critical Metadata QA/QC Workflow (Roles & Responsibilities)

1. Instrument in Need of Servicing
   - Prepare for Shipping
     - Create MAF
     - Enter Purchase Request
     - Log in Refurb Sheet/MAF
     - Create Service Request Form
   - Ship to Vendor
   - Prepare for Integration
     - Create cal files
     - Github
     - Deployment Assignment
     - Update Master Tracking List
   - Prepare for Integration
     - Create MAF
     - Enter Purchase Request
     - Log in Refurb Sheet/MAF
   - Vendor Service
     - Receive from Vendor
     - Inspect and Test
       - Close MAF
       - Complete Instrument Test
       - Update Refurb/MAF log
   - Integration
     - Deployment/Recovery
     - Initialization
       - Checkout First raw file
     - Verification Protocol
       - 2 sets of eyes
     - Pull Request
     - Merge to Github
     - Travis CI Check
     - Release Tag
       - Created Once a Week
     - Metadata Ingestion
       - Cyber Infrastructure Asset Management
       - Github – First Raw File
       - Github – OOI Net
     - Deployment Verification
       - Github – First Raw File
       - Github – OOI Net

Legend
- RCA Data Team
- RCA Engineering Team
- RCA Equipment Manager
- Vendor
- OOI CI
RCA Instrument Calibration Data and Assignment (ICDA)
Verification Workflow

Calibration Data
- OOI Github = RCA Google Drive = APL Refurb Tracking List
  - Yes
  - No
- Calibration Numbers Match
  - Yes
  - No
- Files Use Current Naming Convention
  - Yes
  - No

Issues Addressed and Corrections Applied By a Third Person
  - Yes
  - No

Instrument Assignment
- OOI Github = APL Assignment Sheet
  - Yes
  - No
- Instrument IDs Match
  - Yes
  - No
- Assigned Sites Match
  - Yes
  - No

Recorded on Issue Sheet

Verified by Two People

Verified
RCA 2018 Critical Metadata Entry

- **Calibration sheets:**
  - 77 sensor calibration sheets entered on GitHub
  - 663 parameters includes 6,678 floating point (FP) numbers
  - 72% of FP entries are scripted as of 2016

- **Deployment sheets:**
  - 17 deployment sheets updated on GitHub
  - 96 lines with 12 fields each
  - 1152 fields to enter and verify

- 6,678 calibration coefficients + 1152 deployment fields = 7830 potential sources for data product errors downstream….
RCA Historical (2013-2018) Verification

• **Scope of Issue:**
  - 181 individual instruments with calibrations on GitHub
  - 17 deployment sheets
  - 2013 - 2018 = 5 years of historical data-critical metadata
    7830 fields annually * 5 years = 39150 historical fields to verify

• **10/23/2018 Status**
  All deployment sheets - Checked first-pass
  Calibration data for 68 individual sensors examined
  - Missing calibration files: 46
  - Calibration coefficient errors: 13
  - Calibration coefficient resolution (Seabird): 7
  - File renaming: 32
Sources of Potential Error - Critical Metadata

- **Calibration Coefficients**
  - Example: *Mis-entry of coefficients/filenames on GitHub*
  - Solutions:
    - Scripted entry of calibration coefficients when possible
    - 2i-HITL cross-check of all coefficients on GitHub

- **Instrument Assignments**
  - Example: *Sensor mis-assigned*
  - Solutions:
    - 2i-HITL cross-check of all deployment sheets/cruise info on GitHub
    - Scripted checks of deployment assignment pre-/post-deployment
Sources of Potential Error - Additional

• **Instrument configuration**
  - Example: Incorrect configuration (FW vs. SW mode)
  - Solution:
    - Initialization scripts for each sensor that utilize a database of standardized configurations and commands (in development)

• **Instrument issues**
  - Example: Pump malfunction
  - Solution:
    - Monitor current draw to detect changes in running state of sensor

• **Cyberinfrastructure**
  - Example: Incorrect algorithm
  - Solution:
    - Compare data products with discrete samples, co-located sensors
Sources of Data Errors: Incorrect Algorithm

$pCO_2$ in Water, 2017 Axial Base, Shallow Profiler
Sources of Data Errors: pCO$_2$ in Water
Incorrect Vendor-Provided Algorithm

Cabled Array Profile  
(Original Algorithm)
Discrete Samples
Cabled Array Profile  
(Corrected Algorithm)