Coastal & Global Scale Nodes (CGSN) Data QA/QC Review

November 12, 2019

Sheri N. White, Al Plueddemann, Andrew Reed, Peter Brickley, Stephanie Petillo Jennifer Batryn, Collin Dobson, James Kuo, John Lund, Allen Smith, Rebecca Travis



CGSN Data Team

Data Team Leads

- Sheri N. White Data Team Lead
- **Al Plueddemann** PI/PS
- Andrew Reed Data Science Manager QA/QC*

* Full-Time Data Team members

~2 FTE effort on Data QA/QC in OOI2.0 (Oct 2018-Sept 2019)

Data Team Support

- Peter Brickley Operations
- **Stephanie Petillo** Software
- Jennifer Batryn Instrumentation
- Allen Smith Instrumentation
- Collin Dobson Operations, gliders
- James Kuo Subsurface Moorings
- John Lund Subsurface Moorings
- Rebecca Travis Config. Management



CGSN Data Team Responsibilities

Asset Management

- Entering required metadata (cal, deployment info) into OOINet
- Data Ingestion/Availability
 - Ingesting/re-ingesting data into OOINet
 - Ensuring data flow from platform to shore to science users
- Data QC
 - Identifying/resolving issues with instruments/data
 - Implementing automated and HITL (Human In The Loop) data QC
- Annotation
 - Annotating data in OOINet for instrument/data issues, metadata changes



Asset Management

- Improving asset management processes
 - Double-checking all entered metadata through GitHub
 - Using a dedicated fork and branches for calibrations and deployments
 - Developing automated processes for capturing calibration coefficients
 - To prevent manual entry errors, and speed up the process
 - Documenting Standard Operating Procedures (SOPs)
- Added or Updated ~223 CSV files in PYI of OOI2.0
 - New calibrations, new deployments



CGSN Critical Metadata Review

- Conducted review of all historic Metadata in OOINet
- Scope of work
 - Calibration Coefficients 2749+ calibration files, 30804+ calibration coefficients
 - Platform Deployments
 - Moorings 27 moorings, 162 total deployments
 - Gliders 52 Gliders, 145 total deployments
 - AUVs 2 AUVs, 15 total deployments
 - Assets
 - 1062 Instruments in inventory (sensor_bulk_load)
 - ~300 Mooring control components (eng_bulk_load)



5

CGSN Critical Metadata Review

- Efforts include creation of automated tools for metadata checks
- Tools were necessary due to volume of instrumentation and number of calibration coefficients
 - For example: hundreds of CTDMOs (Conductivity Temperature Depth -Mooring); Large arrays of coefficients for the OPTAA sensors (Optical Attenuation & Absorption)
- Scripts will be used for on-going Asset Management updates
 - To streamline and reduce manual entry errors



CGSN Critical Metadata Review Process

Calibration CSV Review





CGSN Critical Metadata Review Process

Platform Deployment CSV Review





CGSN Critical Metadata Review Status

- Need for metadata review:
 - Non-data impacting: 40% of files have issues related to consistency of naming files
 - Data impacting: 14% of calibration files show errors
- Metadata Review Status:
 - Data impacting: Currently 100% of calibration files reviewed, >90% of updates implemented
 - Non-data impacting: Remaining administrative work for calibration CSVs will be completed in November 2019



Data Ingestion/Availability

- Data ingestions
 - OOI 1.0 known ingestion backlog complete:
 - 25 streams ingested
 - Will perform further ingestions as dictated by Redmine ticket review
 - OOI 2.0 data ingestions:
 - >1500 streams ingested
 - 96% of available ingestions completed
 - Pioneer 12 recovered data ingestions in process
- Data availability
 - Continuous operational review of OOINet, VisualOceans, OMS++
 - Adapting OSU ERDDAP scripts to assess data availability (Gap Analysis)



CGSN Ingestion Process

Telemetered Data





CGSN Ingestion Process

Recovered Data





Data QC

- Operations and Data Monitoring
 - Implemented formal weekly CGSN field operations review
 - Assignment of engineering & data actions, and updating OMS++ infrastructure and instrument alarming
 - Implemented formal weekly CGSN Data Team meetings
 - Assigning/Coordinating data tasks and prioritizations
- Addressing issues via OOI Data Team working groups
 - QARTOD, ADCP (Acoustic Doppler Current Profiler), Communications, Ticketing, Ship-Based Discrete Sampling, etc.
- Working high-nail issues (common failures, Redmine tickets) assigned to CGSN:
 - Helpdesk tickets from Science Users
 - Operations & Hardware tickets generated from CGSN weekly ops reviews



Annotations/Communications

- Annotations are generated as a result of Data Team efforts
 - Monitoring operations of platforms
 - Monitoring data quality of instruments
 - Changes to Critical Metadata that affect data
- Number of annotations created or updated in OOI2.0
 - 170 operational annotations created/updated (24+ more identified to be done)
 - 109 data annotations created/updated (2+ more identified to be done)
- Working on processes for notifying the community about Data Issues
 - Data Communications Working Group has drafted user announcement on Metadata Review, agreed on website location and alternate delivery methods



CGSN Ticketing/Annotation Process



Sources of Issue for Review



CGSN Accomplishments To Date

- Full-time CGSN Data QC hire in November 2018 (Andrew Reed)
- Instituted a weekly CGSN Data Team Meeting
- Attended OOI Data Team Workshop at UW in January 2019
- Completed re-ingestion of backlog from OOI1.0
- Created a common format/spreadsheet for OOI Ship-based Discrete Sampling data across the program
 - Consulted with BCO-DMO (Biological and Chemical Oceanography Data Management Office) on format and community standards
 - Continuing implementation of new format for all past data
 - Enables ingestion of discrete sampling data into OOINet
- Supported development of new CSV format for inductive ADCPs



CGSN Accomplishments To Date

- Successful Asset Management updates and Data Ingestion for:
 - Pioneer 11, 11a, 11b, 11c, 12, 12a, 12b, 12c, 13; Southern Ocean 5; Irminger 5, 6; Papa 6, 7 cruises
- Coordinated a cross-MIO workflow for release of asset management updates to OOINet
- Completed Critical Metadata Review for all data affecting issues
- Supported generation/deployment of single reporting format of Metadata Updates to the OOI user community
- Engaged with IOOS on QARTOD standards
 - Attended IOOS DMAC (Data Management & Communication Subsystem) QARTOD code sprint in Ann Arbor in October 2019



CGSN Plan Going Forward

- Continue weekly OOI & CGSN meetings to identify work and tasking
- Complete administrative activities on Metadata Review
- Evolution of OMS++ system to improve mooring and data monitoring (alerts & alarms, data visualization)
- Support technical refresh activities including data quality review of individual instruments
- Continue to develop and implement automated processes for metadata and data review, support implementation of first QARTOD test in 1Q PYII
- Start hiring process for Data Operations Manager to support day-to-day QA/ QC monitoring, ingestion, and annotation activities



Questions

