OCEAN OBSERVATORIES INITIATIVE

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

October 13, 2020

Sheri N. White, Al Plueddemann, Andrew Reed, Peter Brickley, Stephanie Petillo

Jennifer Batryn, Dan Bogorff, Collin Dobson, James Kuo, John Lund, Allen Smith, Rebecca Travis

OCEANOBSERVATORIES.ORG

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 ¼ FTE effort on Data Activities in PYII

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
- **Dan Bogorff** – Subsurface Moorings
- **Rebecca Travis** – Config. Management

OCEANOBSERVATORIES.ORG

OCEAN OBSERVATORIES INITIATIVE

NSF

CGSN Data Team Responsibilities

- **Asset/Metadata Management**
  - Entering required metadata (calibration, deployment info) into OOINet
  - Organizing, cataloging and posting shipboard and water sampling data
- **Data Ingestion/Availability**
  - Ingesting/re-ingesting data into OOINet
  - Ensuring data flow from platform to shore to science users
  - Supporting OOI data distribution to other data providers
- **Data QC**
  - Identifying/resolving issues with instruments/data
  - Implementing automated and HITL (Human In The Loop) data QC
- **Annotation/Communications**
  - Annotating data in OOINet for instrument/data issues, metadata changes

CGSN Asset Management

- Continuing to use our improved asset management processes
  - Double-review all updates/additions in WHOIGit/ooicgsn-asset-management fork
    - Using automated code when possible to generate calibration CSVs
  - Push to OceanObservatories/asset-management repository

- In PYII of OOI 2.0 – **891** files added or updated
  - New instrument calibrations
  - Metadata updates
  - Cruises – Pioneer 13, 13a, 13b; Southern Ocean 6; Irminger 7; Pioneer 14, 14a, 14b, 14c; Papa 7a and other glider recoveries

Roundabout Asset & Metadata Tracking

CGSN supported development of **RDB** as Asset & Metadata Tracking Tool

- **Roundabout CI** Status:

  In PYII

  - Completed definition of requirements, supported by CI and all MIOs.
  - 42% of overall development tasks completed as planned.

  In PYIII

  - ECR-489, approved by NSF, re-allocated PYII underruns to Roundabout CI PYIII budget.
  - Remaining 58% of development tasks to be completed by end of Mar 2021, followed by final implementation support.

CGSN Metadata Review

Calibration CSVs

- **2802** calibration files reviewed (over **30,000** individual coefficients)
- **14%** data affecting changes

Deploy CSVs

- **151** Mooring deployments (over **3,000** lines of instruments/components)
- **178** Glider/AUV deployments
- **7%** were data affecting changes

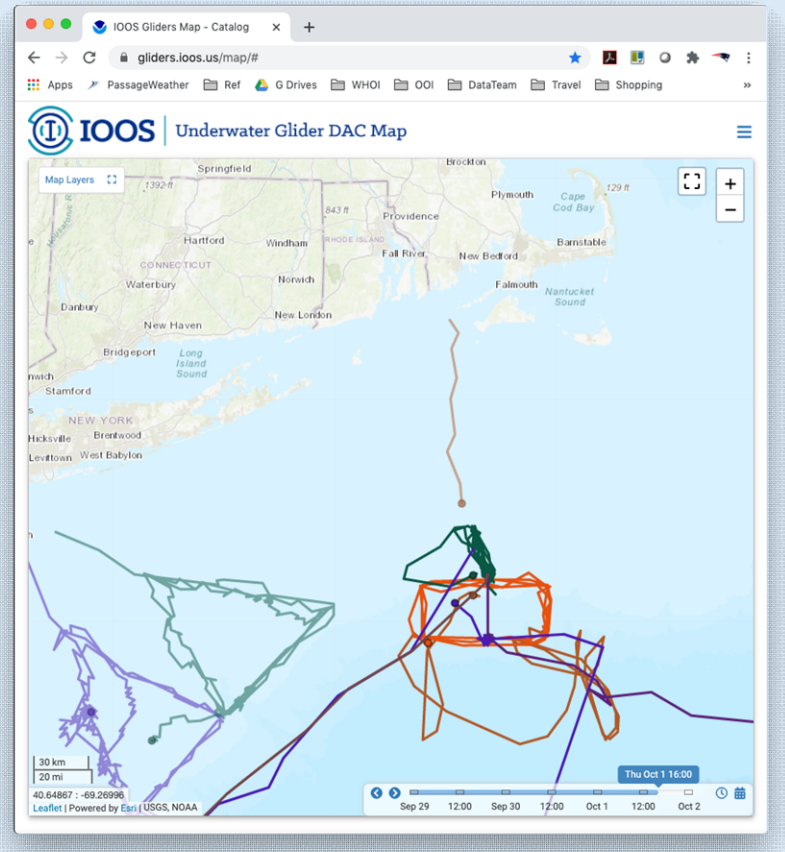
- All data reviewed and updated
- Data effecting changes presented in the OOI Metadata Widget

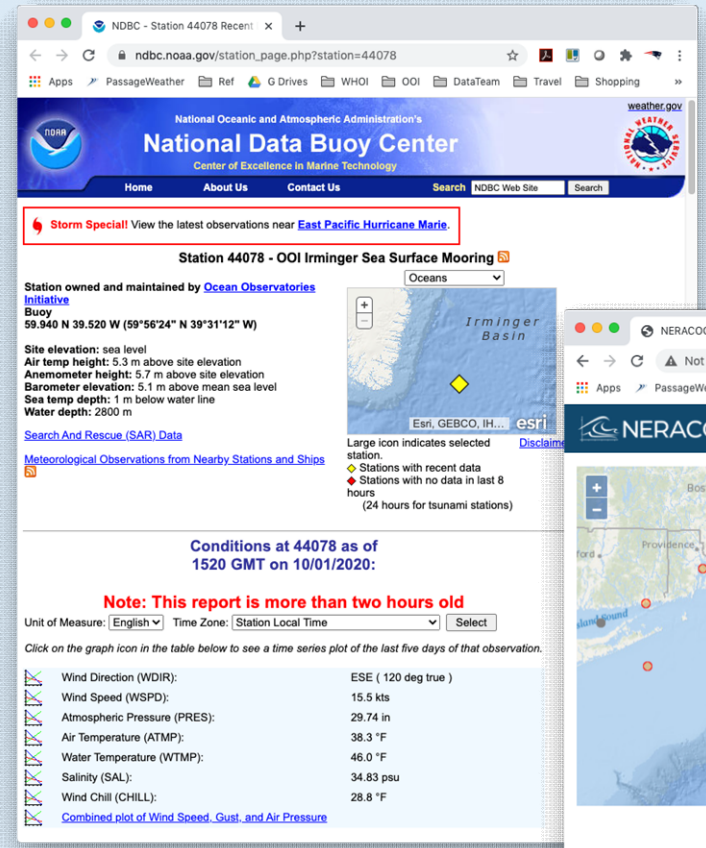
- Final step is clean up of bulk load and CruiseInformation CSVs

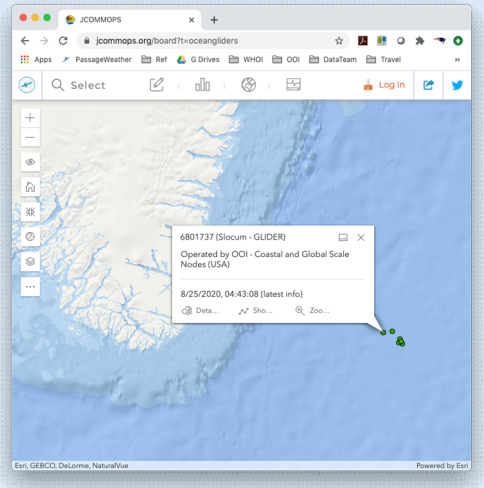
CGSN Data Ingestions/Availability

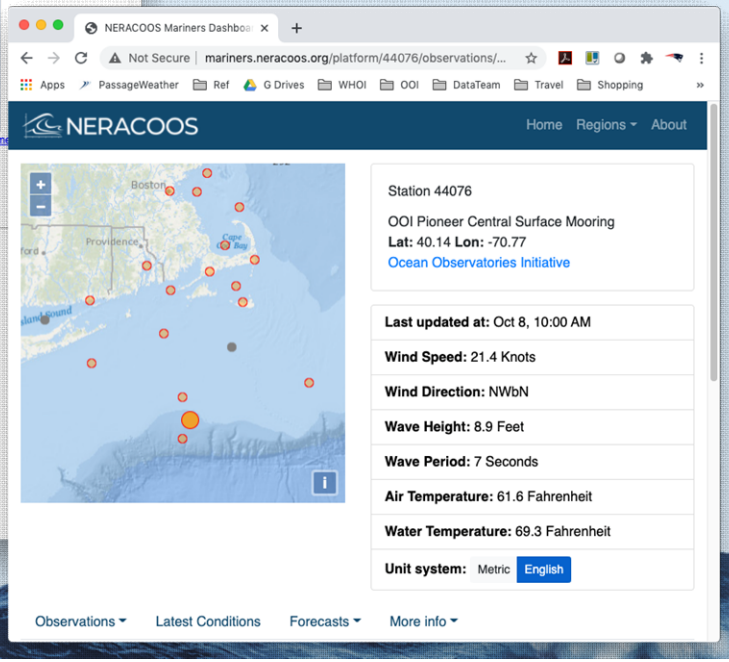
- **Data ingestions**
  - OOI 2.0 PYII data ingestions:
    - **~1,500** new ingestions and **~3,000** re-ingestions to address issues and missing data
- **Data availability/distribution**
  - Continuous operational review of data within OOINet, OMS++
  - Distributing glider data to **IOOS Glider DAC**
    - **165** Pioneer and Global deployments identified for the Glider DAC
    - Active deployments have been pushed since May 2020
    - Working backwards for older deployments (**~20** submitted to date)
    - Also in **OceanOPS** (formerly **JCOMMOPS**) part of **GOOS**
  - Distributing Surface Mooring METBK and WAVSS data to **NDBC**
    - Also in **NERACOOS**

CGSN Data Ingestions/Availability









CGSN Data QC

Operations and Data Monitoring

- Continuing formal weekly CGSN operations review & CGSN Data Team meetings
- Implemented CGSN Incident Tracking document
  - Track infrastructure & instrument issues to closure, generate statistics on types/sources of issues

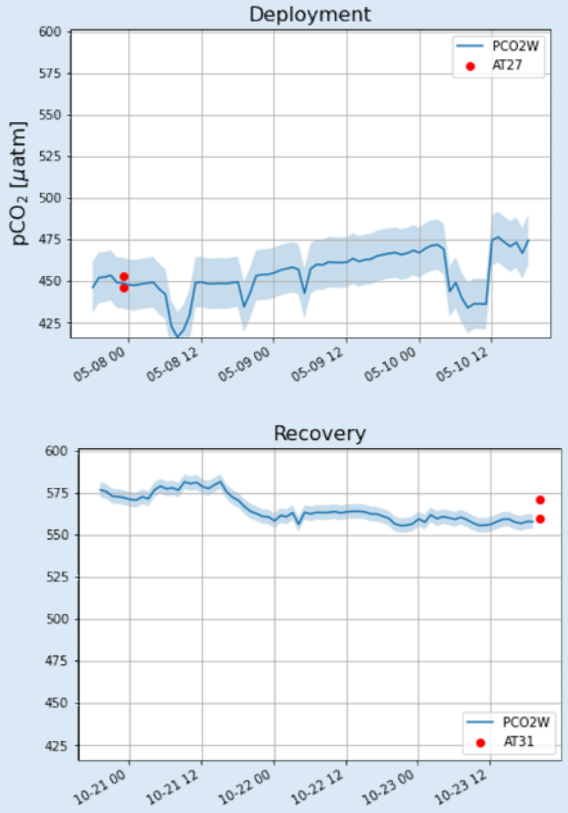
- Working high-nail issues – common failures, Redmine tickets assigned to CGSN:
  - Helpdesk tickets from Science Users (**41** CG Redmine tickets closed out in PYII)
  - Operations & Hardware tickets generated internally (**62** CG tickets closed out in PYII)

- Implemented regular CGSN Data QC meetings for discussing tasking and prioritization of deeper data dive efforts

CGSN Data QC

- Assessing data quality to support Instrument Tech Refresh Efforts
  - Supported PHSEN analysis (started by EA)
    - Evaluated data quality via gross range test and noisiness
    - Identified instrument failure modes and vendor issues
    - Conducted a deep dive and bottle data comparison at Irminger
  - Begun implementation of same analysis for PCO2W
    - Deep dive and bottle data comparison
    - Downloaded all PCO2W data to identify data gaps and evaluate data quality

CP01CNSM MFN PCO2W
Deployment 3



CGSN Data QC

QARTOD

- **CTD instrument-class** – supporting implementation by QARTOD Working Group
  - **Gross Range Test (GRT): 100% of tasks complete**, In production.
  - **Climatology Test (CT): 100% of tasks complete**, In production. MIOs will perform one final test to ensure all actions were completed during handover from Raytheon.
  - **Timing & Gap Tests:** Agreed on OOI specific philosophy for Timing & Gap tests. Software development on hold pending completion of priority CI tasks.
    - **Timing Test** would be applied on an instrument level to provide operator with indication of functioning instruments; verifies receipt of data at expected intervals.
    - **Gap Test** would be applied to telemetered & recovered data to locate gaps in data, test sampling rate.

- Next focus is on increasing the number of instruments available for GRT and CT testing. Proposed prioritization of instruments, to be reviewed by PI/PS, will mirror the Tech Refresh priority list.

CGSN Data QC

OOI 1.0 Data Review

- **1497** issues identified with CGSN platforms
- All issues to be categorized as:
  - **Known – Fixed**
    - A number of issues were already resolved by the Critical Metadata Review and other ongoing work of the data team
  - **Known – In Progress**
    - A number of issues have been identified in Redmine and are in the process of being addressed (e.g., pressure issues and glider data issues)
  - **Unknown – To Be Addressed**
    - Some items were not yet identified and will be addressed
- We are addressing these issues as we come to them through the course of our work, and as we have time to delve in to new issues.

CGSN Annotations/Communications

- Annotations are generated as a result of Data Team efforts
  - Monitoring operations of platforms, and data quality of instruments
  - Redmine tickets and 1.0 Data Review

- Number of annotations created or updated in OOI 2.0 PYII
  - **146** operational annotations created/updated
  - **227** data annotations created/updated

- Communicate data changes via various other methods as well
  - Notices on the Ocean Observatories Initiative website
  - Data-affecting changes are added to the **OOI Metadata Widget**
  - Raw Data Repository is annotated using `00ReadMe.txt` files

CGSN PYII Accomplishments

- Attended OOI Data Team Workshop at WHOI in January 2020
- Supported OOI booth at Ocean Sciences 2020 in February 2020

- Successful Asset Management updates and Data Ingestion for:
  - Pioneer 13, 13a, 13b; Southern Ocean 6; Irminger 7; Pioneer 14, 14a, 14b, 14c; Papa 7a and other glider recoveries
- Completed Critical Metadata Review for all data affecting issues
  - Consolidated CGSN metadata changes for implementation of the OOI Metadata Widget

- Distributed data to other repositories
  - Submitting CGSN Glider data to IOOS Glider DAC
  - Submitting CGSN Surface Mooring data to NDBC

CGSN PYII Accomplishments, cont.

- Participated in cross-MIO Working Groups to address specific issues
  - **QARTOD** – Supported implementation of QARTOD tests for CTDs
  - **Keryx** – Supported implementation OOI Data Explorer by Axiom
  - **RDB** – Generated cross-MIO requirements and began implementation of RDB as the OOI Asset & Metadata Tracking system
  - **Pre-Load** – Supported review and update of parameter units, definitions and terminology
  - **Raw Data Repo Re-organization** – Determined the scope of work and mechanism to track progress
  - **Ticketing Re-organization** – Restructuring Redmine for use throughout OOI
  - **DPS** – Supported updates to Data Product Specifications (PCO2W, ADCP, PAR)
  - **ADCP** – Supported development and implementation of new calibration CSV format for inductive ADCPs, and addressed other issues [Work completed and WG disbanded]
  - **ZPLSC/G** – Supporting development of ZPLSC/G echogram generation code and other data issues
  - **Pressure** – Supporting solution to Profiler and other pressure issues

CGSN PYII Accomplishments, cont.

- Reviewed all shipboard and water sampling data, and updated for standard filenames
  - Identified and posted missing data/files; reviewed/corrected digitized CTD logs
  - Discrete Sampling Summary Spreadsheets completed for Irminger 1-6

- Redmine Issues
  - **125** CGSN-related Redmine tickets closed out in PYII across all Redmine projects
  - Supported efforts to correct Lat/Lon issues with Glider data

- Supported re-organization of OOI GitHub repositories into a common organization

CGSN Path Forward

- Continue weekly OOI & CGSN meetings to identify work and prioritize tasking
- Contribute to the evolution of OMS++ system to improve mooring and data monitoring (alerts & alarms, data visualization)
- Support Roundabout development for Asset & Metadata Tracking
- Increase number of historical Glider deployments in the IOOS Glider DAC
- Expand Surface Mooring data delivery to NDBC (more variables)
- Support implementation of QARTOD for prioritized instruments
- Continue to address science user Helpdesk questions
- Deliver an initial set of discrete water sampling Summary Spreadsheets per the approved common format to science users
- Support Tech Refresh with data analysis for prioritized instruments
- Add 1 FTE to support data operations/monitoring and OMS++

OCEAN
OBSERVATORIES
INITIATIVE

Questions?

OCEANOBSERVATORIES.ORG