



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

May 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



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 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





CGSN Data Team Responsibilities

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



CGSN Asset Management

- Continuing to use our improved asset management processes
 - Double-checking all metadata through dedicated forks and branches in GitHub
 - Using and developing automated processes for capturing calibration coefficients
- Added or updated **947+** CSV files in PYII of OOI 2.0
 - New instrument calibrations, new deployments
 - Metadata updates
- Documenting Standard Operating Procedures (SOPs)
- Supporting development of **Roundabout** as Asset/Metadata Management Tool
Selected during the CI Analysis of Alternatives (AoA) as the tool to deliver Metadata information to CI
 - The MIOs and CI have finalized requirements, developed goals, and begun to implement Roundabout updates to support this effort



CGSN Metadata Review

- **Need for metadata review**
 - Non-data impacting: **40%** of files had issues related to consistency of naming files
 - Data impacting: **14%** of calibration files showed errors
- **Metadata Review Status**
 - Currently **100%** of calibration files reviewed and implemented
 - Currently **100%** of deployments reviewed, **60%** implemented
 - Planned completion in **May 2020**
- Data effecting changes presented in the OOI Metadata Widget



CGSN Data Ingestions/Availability

- **Data ingestions**

- OOI 2.0 PYII data ingestions:
 - **989** streams ingested (new and re-ingestions to address Redmine tickets)
 - **100%** of available ingestions completed

- **Data availability/distribution**

- Continuous operational review of OOI Net, OMS++; and gap analyses with OSU scripts
- Working on distributing data to **IOOS Glider DAC**
 - **151** Pioneer and Global deployments identified for the Glider DAC
 - **3** of **4** currently-deployed Pioneer gliders in the Glider DAC, and 55 early deployments submitted by Rutgers
- Working with **NDBC** to distribute Surface Mooring METBK and WAVSS data
 - WMO numbers obtained for the Pioneer and Irminger Surface Moorings
 - Processing code is tested, and we are 90% ready to start pushing data



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 (**26** CG Redmine tickets closed out in PYII)
 - Operations & Hardware tickets generated internally (**29** CG tickets closed out in PYII)
 - Addressed ADCP issues related to time stamps, and necessary metadata for telemetered inductive ADCP data, etc.

CGSN Data QC

- Assessing data quality to support Instrument Tech Refresh Efforts
 - Supporting PHSEN analysis started by EA
 - Identify data gaps and re-ingest missing data
 - Evaluate data quality via gross range test
 - Identify instrument failure modes and vendor issues
 - Begun implementation of same analysis for PCO2W
- Addressing cross-MIO issues via OOI Data Team working groups
 - **QARTOD, Roundabout, Keryx**, ADCP (Acoustic Doppler Current Profiler), Communications, Ticketing re-org, Ship-Based Discrete Sampling, Data Product Specifications (DPS), Pre-Load, ZPLSC/G, Pressure



CGSN Data QC

- **QARTOD**

- MIO responsibilities include:
 - Selecting instrument-classes for QARTOD development
 - Selecting applicable QARTOD tests to be developed for specific instrument-class
 - Selecting test data streams to support development
 - Defining and supplying instrument test data (ranges, binning, etc.)
 - Running example queries to test code and providing feedback
- **CTD instrument-class** – supporting implementation by QARTOD Working Group
 - Gross Range, Climatology, Timing tests selected for development
 - **Gross Range Test: 88% of tasks completed**, final task going to production June 2nd
 - **Climatology Test: 50% of tasks complete**, next finish review of test data methodology
 - **Timing Test:** Reviewing test philosophy and generating goals specific to OOI



CGSN 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 OOI 2.0 PYII
 - **146** operational annotations created/updated
 - **77** data annotations created/updated (**8** more identified to be done)
- Data-affecting changes are added to the **OOI Metadata Widget**
- Working on process to annotate Raw Data Repository and other data locations
 - Currently using `00ReadMe.txt` files in the Raw Data Repository



CGSN Accomplishments

- Attended OOI Data Team Workshop at WHOI in January 2020
- Successful Asset Management updates and Data Ingestion for:
 - Pioneer 13a, 13b; Southern Ocean 6; and Pioneer and Irminger glider recoveries
 - Initial Asset Management updates for Pioneer 14, Irminger 7, and Papa 8 cruises
- Completed Critical Metadata Review for all data affecting issues
 - Consolidated CGSN metadata changes for implementation of the OOI Metadata Widget
- Reviewed all shipboard and water sampling data, and updated for standard filenames
 - Identified and posted missing data/files; reviewed/corrected digitized CTD logs
- Supported development and implementation of new calibration CSV format for inductive ADCPs – driven by efforts of the ADCP Working Group
- Supported re-organization of OOI GitHub repositories
- Commenced process to submit CGSN glider data to IOOS Glider DAC
- **55** CGSN-related Redmine tickets closed out in PY11



CGSN Path Forward

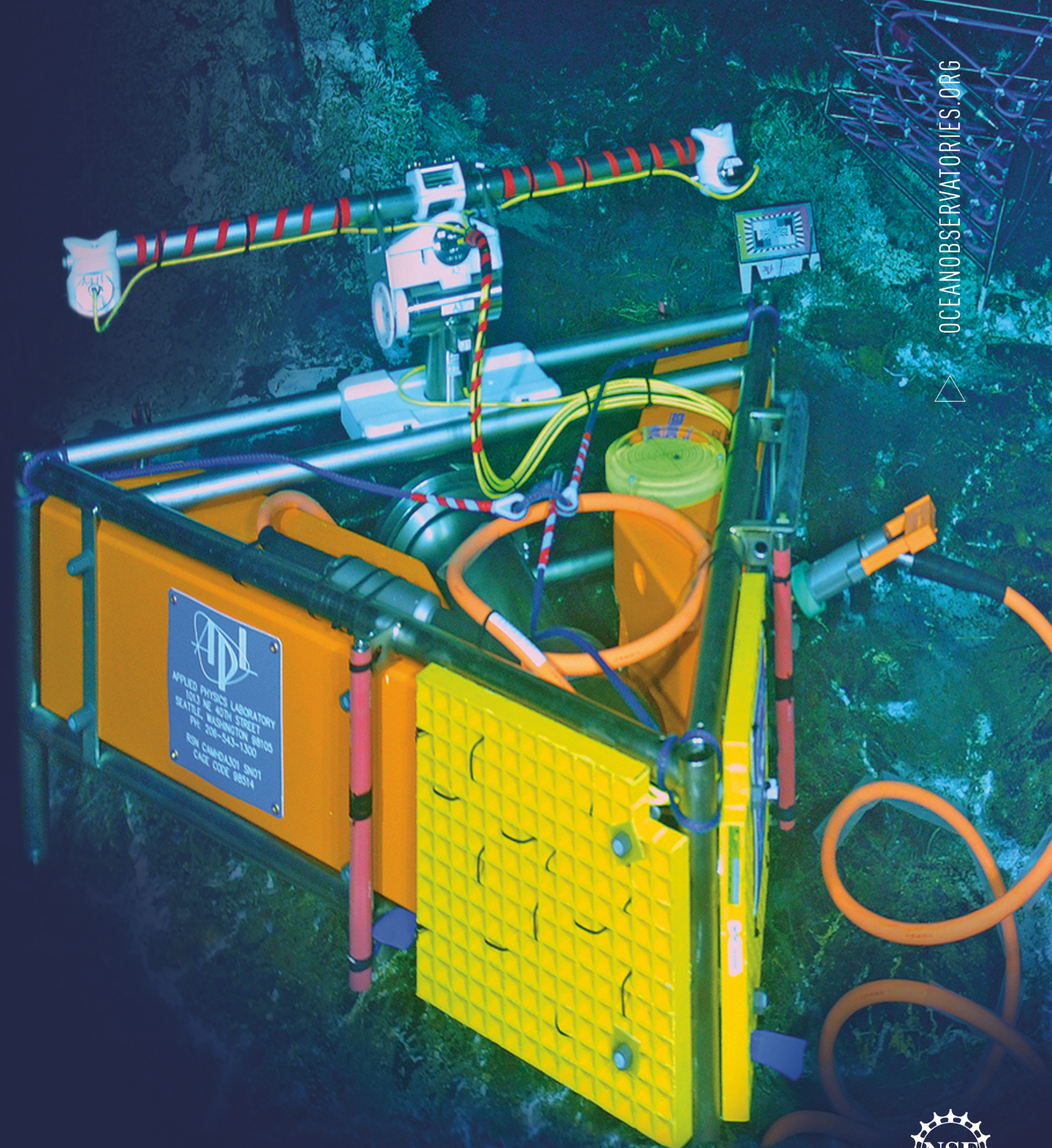
- Continue weekly OOI & CGSN meetings to identify work and prioritize tasking
- Continue evolution of OMS++ system to improve mooring and data monitoring (alerts & alarms, data visualization)
- Continue to support implementation of first QARTOD tests
- Continue to support Roundabout development of Asset Management and CI Metadata processes
- Continue glider data submittals to IOOS Glider DAC
- Begin surface mooring data submittals to NDBC
- Support data analysis for Instrument Tech Refresh efforts
- Continue to post water sampling data and generate Summary Spreadsheets per the approved common format





OCEAN
OBSERVATORIES
INITIATIVE

Questions?



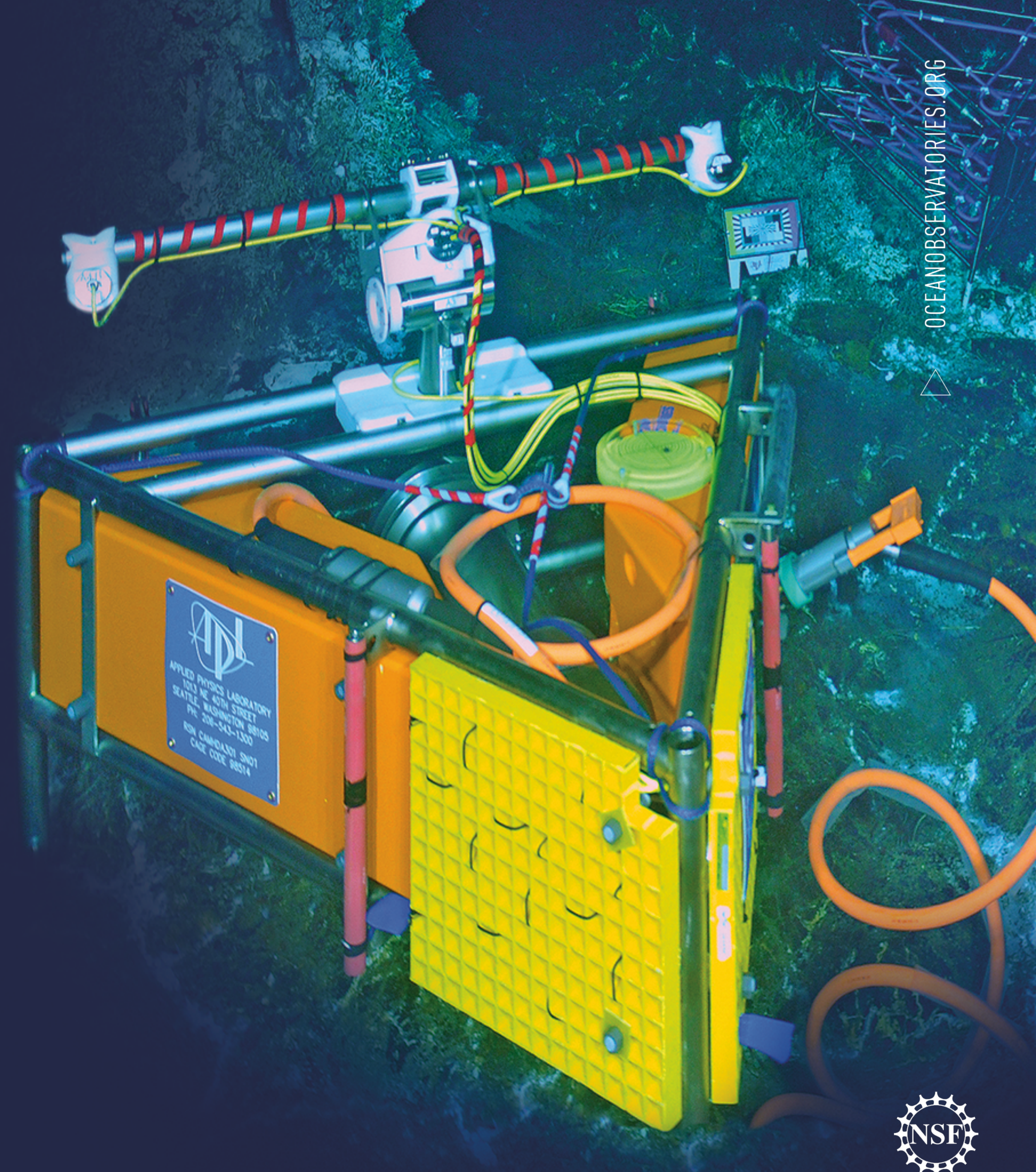
OCEANOBSERVATORIES.ORG





OCEAN
OBSERVATORIES
INITIATIVE

Back-up Slides



OCEANOBSERVATORIES.ORG





CGSN Asset Management

- **Roundabout**

Roundabout is a WHOI developed tool to support:

- Asset & Material Tracking
- Storage of Calibration Metadata
- Building and Storage of Deployment Configurations & Information

Roundabout was selected during the CI Analysis of Alternatives (AoA) as the tool to deliver Metadata information to CI. The MIOs and CI have finalized requirements, developed goals, and begun to implement Roundabout updates to support this effort