

# Regional Cabled Array (RCA) Data QA/QC: PY1 Summary and Projected PY2 Activities

Wendi Ruef, Mike Vardaro, Orest Kawka, and RCA Team

OOI Regional Cabled Array

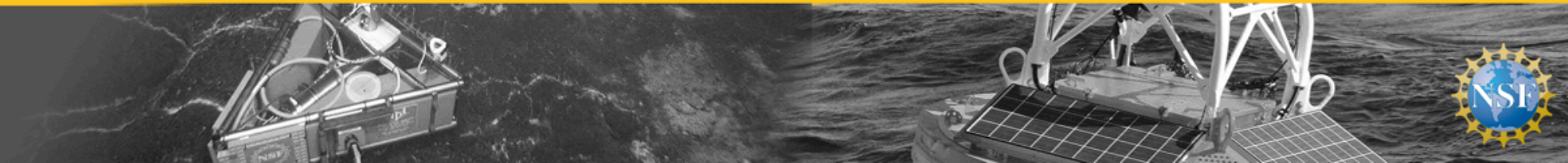
School of Oceanography

University of Washington

Seattle, WA 98195

OOIFB and DDCI Committee Meeting

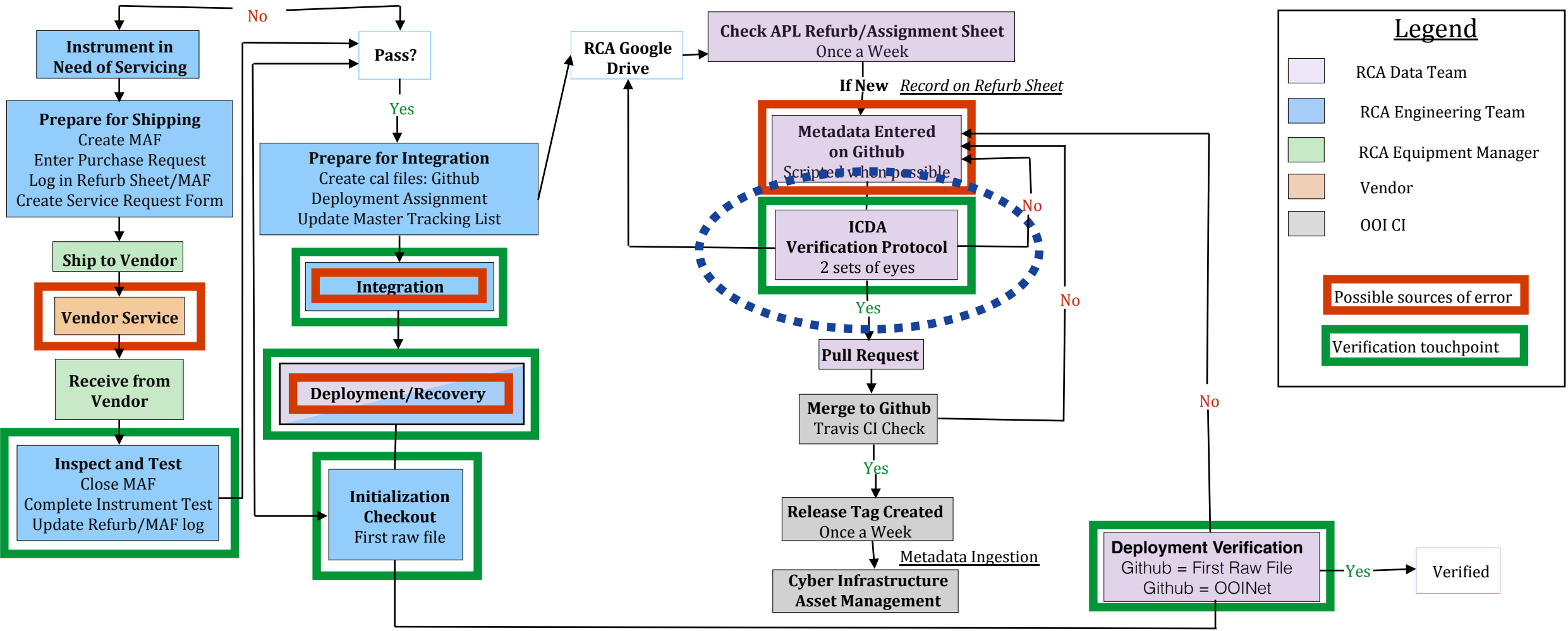
Nov 12-13, 2019



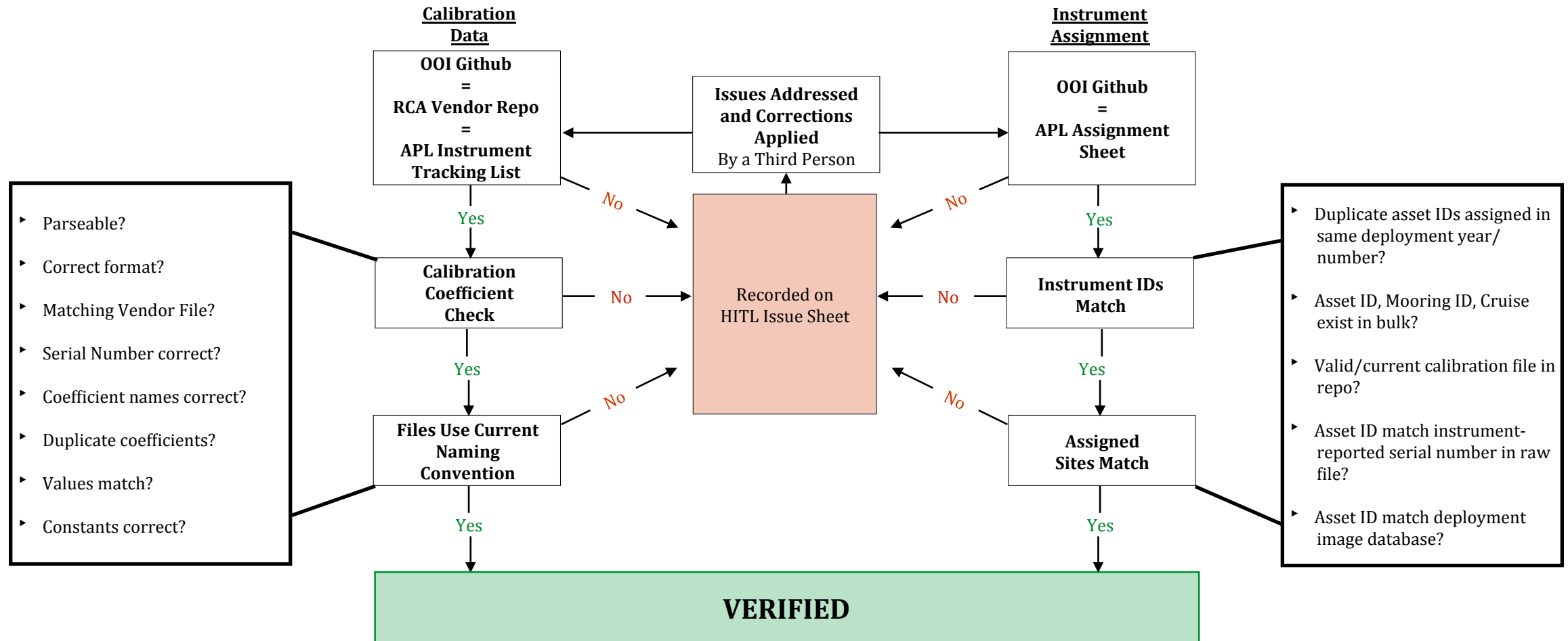
# RCA Data QA/QC Prioritization for PY 1

- ▶ **Oct 1, 2018 – RCA operational for 4 years (2014-2018)**
  - RCA data available via OOI Net and viewable and downloadable
  - Rutgers Data Team had created and verified RCA data product baseline
  - Existing data gaps required “playback” from raw data – tool not available to MIO
  - Accuracy/validity of data end products highly dependent on “Critical Metadata” in CI
    - Calibration information
    - Instrument assignment mapping
  - High Priority – Providing users confidence in quality of available data
- ▶ **RCA PY 1 – Prioritized Critical Metadata Verification**

# RCA Instrument and Critical Metadata (CrM) QA/QC Workflow (R&R)



# RCA Instrument Calibration Data and Assignment (ICDA) Workflow



# RCA PY 1 Priority: Historical Critical Metadata Check

## ► CRITICAL METADATA (CrM) AUDIT RESULTS:

- Calibration Files:
  - 133 data-affecting modifications (98 missing files, 32 incorrect coefficients, 3 file renames)
  - “data-affecting”: discrepancies < 7 significant figures, discrepancies to coefficients associated with active deployment instances, file dates overlapping active deployments
- Deployment Files:
  - 14 data-affecting modifications
  - “data-affecting”: discrepancies in asset IDs, deployment dates, location assignments

# Critical Metadata Check Lessons Learned

- ▶ **Missing files accounted for majority of calibration errors**
  - ◆ Curated repo of vendor calibration documentation
  - ◆ Verification flag for deployment instances without valid or current calibration files
- ▶ **Scripted entry does not eliminate all errors**
  - Inconsistencies in coefficient resolution and file types
    - ◆ Verify highest resolution available used for coefficient entry
  - Typos and erroneous coefficients not always obvious
    - ◆ Validate vendor files for consistency in dates, headers, ranges of coefficients
- ▶ **Need for consistent gold Standards**
  - ◆ Pre/post deployment image database cross-referenced with deployment logs
  - ◆ Mining sensor information in raw file during QCT testing and post-deployment





# RCA PY1 Data QA/QC Summary

## ► MILESTONES ACHIEVED:

- Developed RCA data QA/QC workflow and processes/procedures for CrM
- Completed audit of RCA historical non-Tier 1 CrM
- Corrected/re-verified 95% of RCA historical CrM for 2014-2018
- Developed/implemented QC scripts to automate routine checks of CrM
- Created Shipboard Discrete Sample Summary template
  - MIOs collaborated cross-program for creation of common format
- Created, uploaded, and verified 105 instrument updates for 2019 RCA Maintenance Cruise
- Created prototype GUI widget for user access to OOI CrM correction database
- Assumed responsibility for maintenance/monitoring of RCA Port Agents and Drivers

# RCA PY1 Data QA/QC Summary (cont)

## ► OTHER /ONGOING:

- OOI Data Management Working Group and subgroups: QARTOD, Redmine, Comms, AoA
- Select sensor deep dives, including O<sub>2</sub> optodes, broadband hydrophones, pCO<sub>2</sub> cal ranges
- Monitoring data flow from instruments through to OOI Net and Raw Servers: Availability
- Active investigation/resolution of data-related Redmine Tickets, including HelpDesk
- Entrained volunteer/paid undergraduate students in RCA data projects and QA/QC

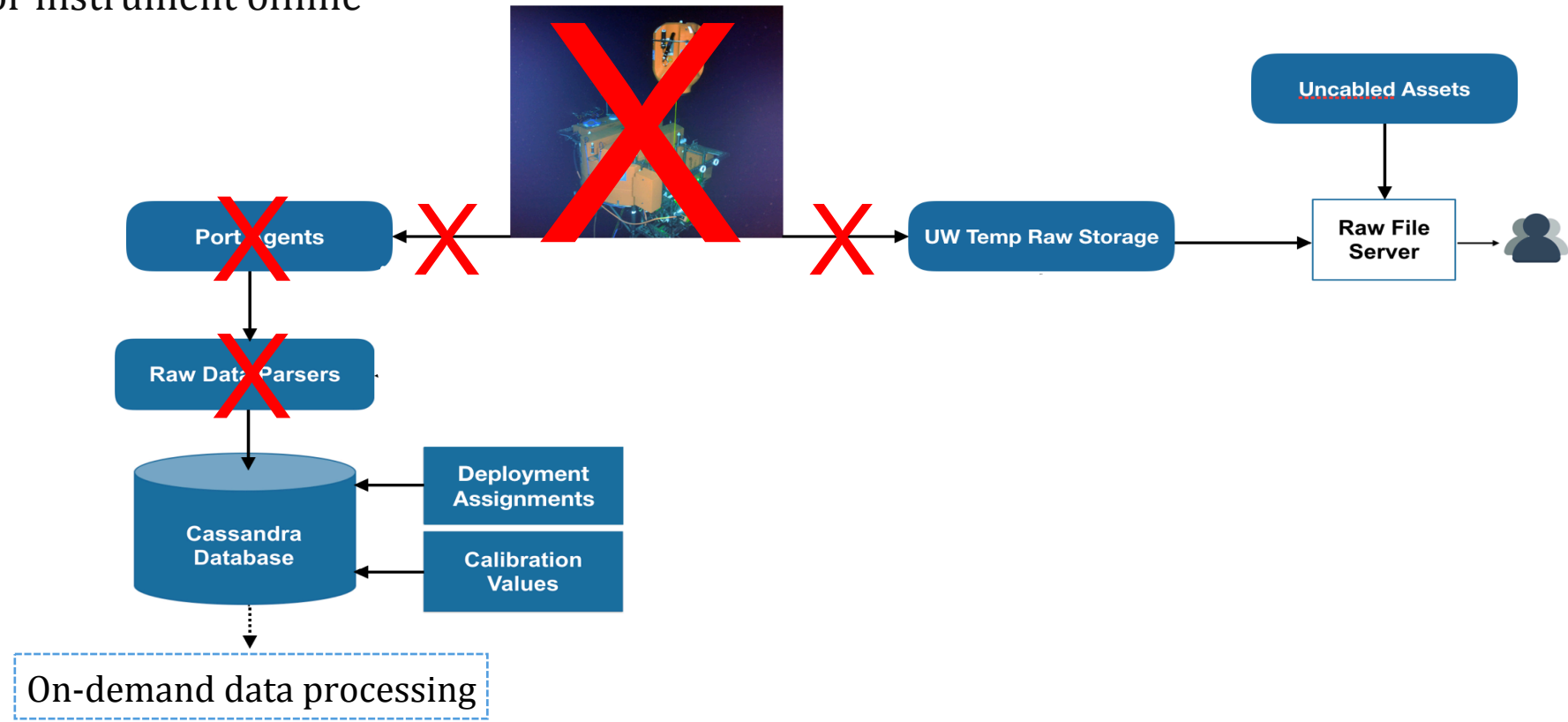




# Data Delivery and Availability

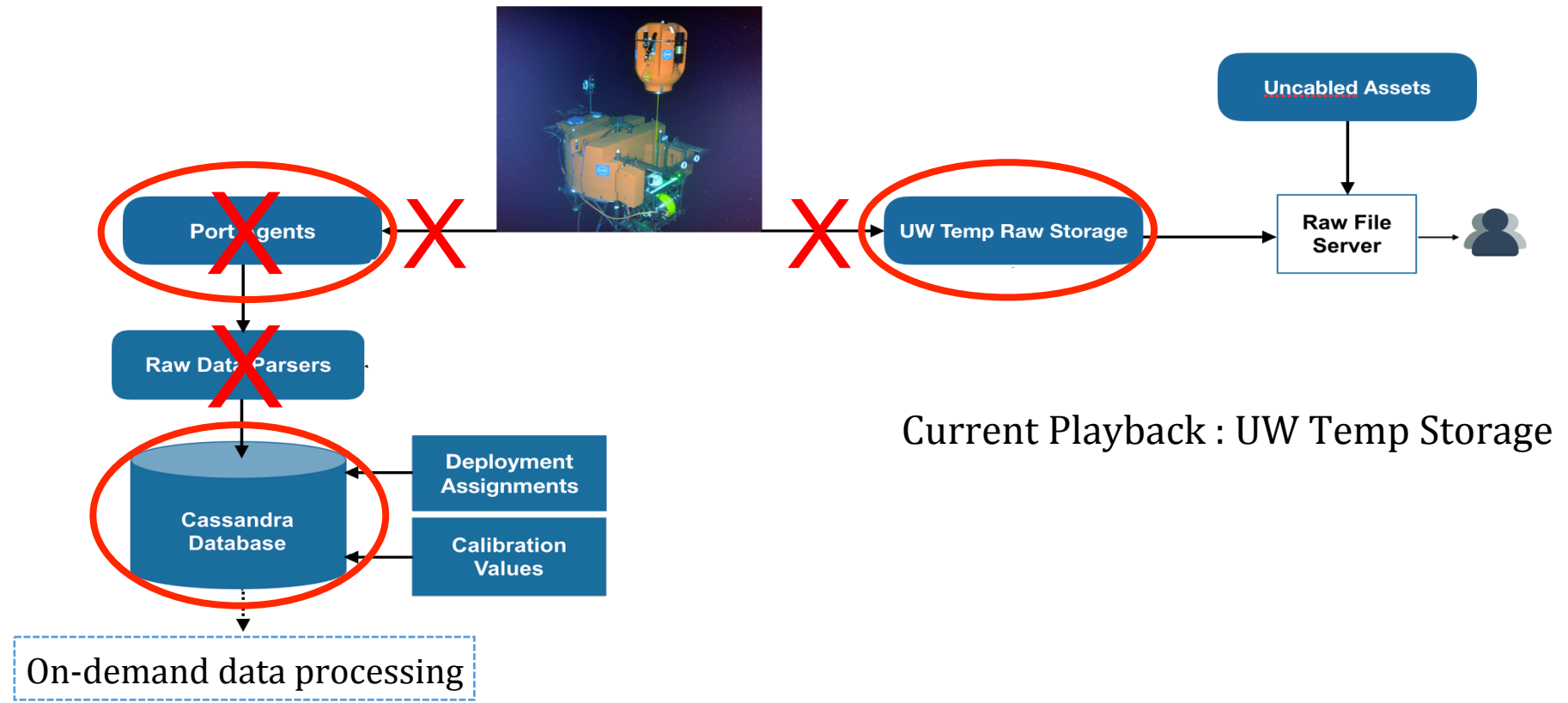
## ► Causes of missing streamed (cabled) data

- Network outages
- Port Agent / Parser errors
- RCA or instrument offline



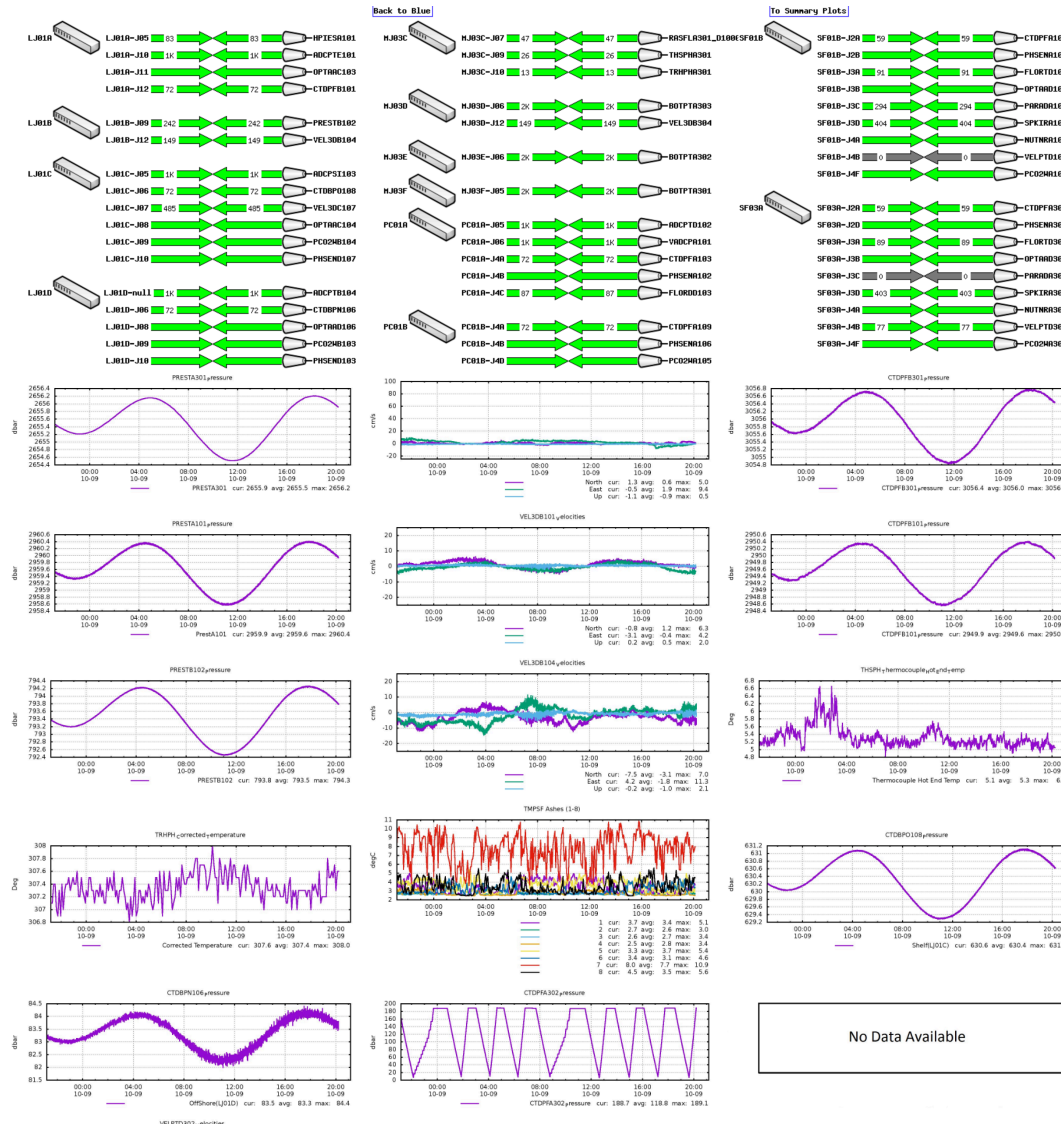
# Data Delivery and Availability

- ▶ **Resolution of missing streamed (cabled) data**
  - Cabled Playback
  - Availability analysis requires comparison of raw data archives with Cassandra database



# RCA Data Monitoring Dashboards

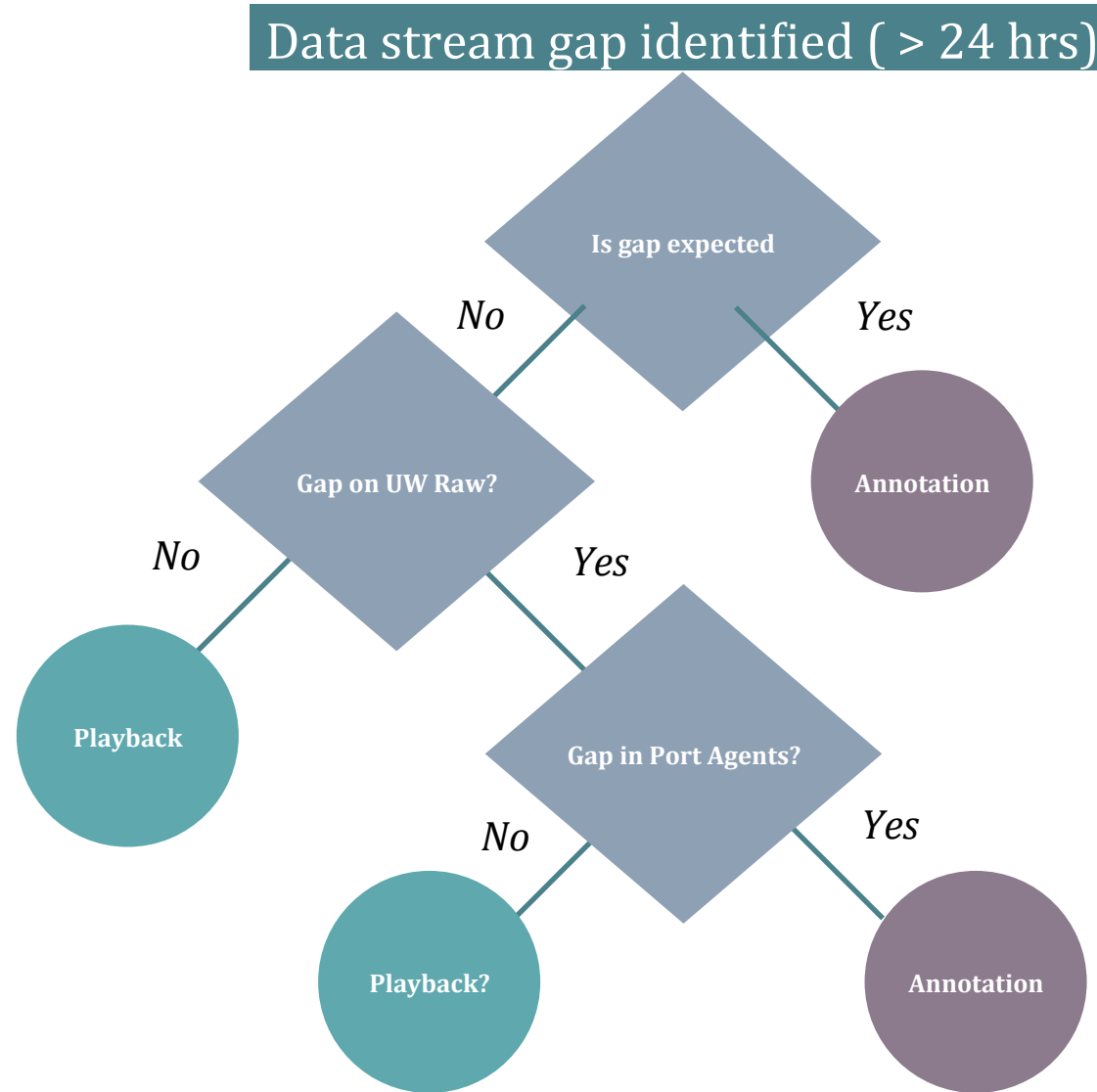
**Note:** Table updates hourly. If viewing near the top of the hour, please allow time for the list to populate.



Instrument	Status	Data End Time	Offset
CE04OSBP-LJ01C-05-ADCPSI103		<b>M2M Error</b>	<b>M2M Error</b>
RS01SUM2-MJ01B-12-ADCPSK101		<b>M2M Error</b>	<b>M2M Error</b>
CE02SHBP-LJ01D-05-ADCPTB104		<b>M2M Error</b>	<b>M2M Error</b>
RS01SBPS-PC01A-05-ADCPTD102		2019-10-09 20:01:08 UTC	3.3 sec
RS01SLBS-LJ01A-10-ADCPTE101		2019-10-09 20:01:23 UTC	2.3 sec
RS03AXBS-LJ03A-10-ADCPTE303		2019-10-09 20:01:43 UTC	1.0 sec
RS03CCAL-MJ03F-05-BOTPTA301		2019-10-09 20:02:01 UTC	0.1 sec
RS03ECAL-MJ03E-06-BOTPTA302		2019-10-09 20:02:16 UTC	4.7 sec
RS03INT2-MJ03D-06-BOTPTA303		2019-10-09 20:02:47 UTC	0.4 sec
RS03ASHS-MJ03B-09-BOTPTA304		2019-10-09 20:03:08 UTC	2.3 sec
CE02SHBP-LJ01D-06-CTDBPN106		2019-10-09 20:03:21 UTC	1.3 sec
CE04OSBP-LJ01C-06-CTDBPO108		2019-10-09 20:03:45 UTC	1.0 sec
RS01SBPS-SF01A-2A-CTDPFA102	OFFLINE	Operationally Offline	---
RS01SBPS-PC01A-4A-CTDPFA103		2019-10-09 20:04:18 UTC	2.2 sec
CE04OSPS-SF01B-2A-CTDPFA107		2019-10-09 20:04:43 UTC	2.6 sec
CE04OSPS-PC01B-4A-CTDPFA109		2019-10-09 20:05:02 UTC	3.1 sec
RS03AXPS-SF03A-2A-CTDPFA302		2019-10-09 20:05:38 UTC	1.2 sec
RS03AXPS-PC03A-4A-CTDPFA303		2019-10-09 20:06:07 UTC	3.5 sec
RS01SLBS-LJ01A-12-CTDPFB101		2019-10-09 20:06:30 UTC	2.5 sec
RS03AXBS-LJ03A-12-CTDPFB301		2019-10-09 20:06:58 UTC	4.2 sec
RS03ASHS-MJ03B-10-CTDPFB304		2019-10-09 20:07:20 UTC	1.1 sec
RS01SBPS-PC01A-4C-FLORDD103		2019-10-09 20:06:21 UTC	1.3 min
RS01SBPS-SF01A-3A-FLORTD101	OFFLINE	Operationally Offline	---
CE04OSPS-SF01B-3A-FLORTD104	OFFLINE	Operationally Offline	---

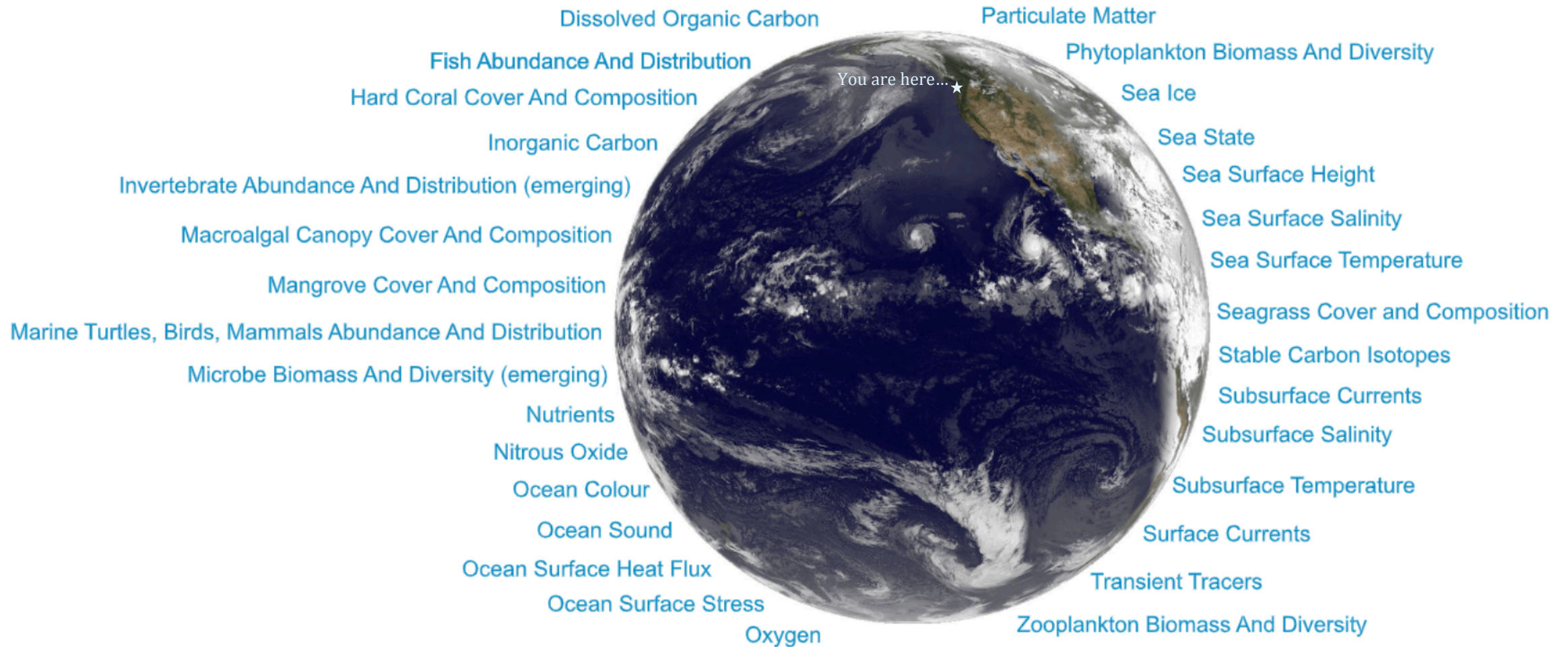
# Data Availability

- ▶ Helpdesk
- ▶ VisualOcean
- ▶ OOI Net





# In case you want to see the essential ocean variables...



# RCA PY2 Data QA/QC

## ► NEAR-TERM PRIORITIES (PY2 Q1):

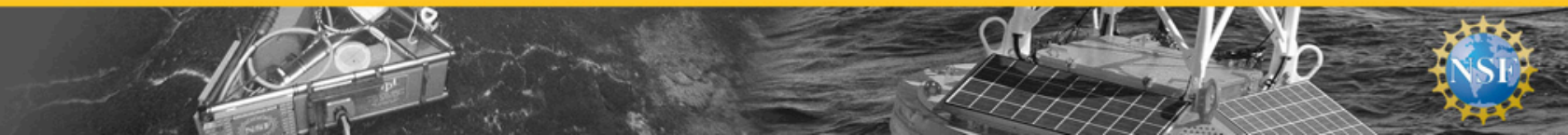
- Data annotations and user communication for audit of historic CrM
- Upload to Alfresco of RCA Discrete Water Sample Data from 2014-2019 Field Verifications  
2015-2019 Completed ✓
- Investigation/resolution of DOSTA Calibration and Algorithm Issues
- Investigation of data flow interruptions: BOTPT and ZPLSC
- Verification of seismic instrument (TIER 1) CrM at IRIS
- Verification/ingestion of CrM for Deep Profilers and RCA Platforms (engineering data)



# RCA PY2 Data QA/QC (cont)

## ► MID-TERM PRIORITIES (PY2 Q2):

- Development of workflow for assessing RCA data availability and resolution thereof: Cabled Playback
- Apply historical RCA data availability check and resolution thereof: Cabled Playback  
**Start: Oct 1, 2018 (PY1) GAPS > 24 hours**
- Creation of transfer functions and calibration data repository for HYDBB data
- Development of RCA-specific HITL data QA/QC workflow using QARTOD results and discrete samples
- Deployment image database
- Updates/testing of instrument drivers/parsers for PAR, THSPH



# RCA PY2 Data QA/QC (cont)

## ▶ LONG-TERM PRIORITIES (PY2 Q3-Q4):

- Completion of historical data availability check with Cabled Playback

**TBD: Gap Size Tolerance, “Look Back”, Parameter Breadth**

- Application of RCA-specific HITL data QA/QC workflow; sensor deep dives as needed
- Asset management upload and verification for 2020 deployments
- New PI instrument data flow verification
- Participation in RCA O&M Cruise (VISIONS'20)



# RCA PY2 Data QA/QC (cont)

## ▶ ONGOING ACTIVITIES (PY2):

- Assist with design and implementation of QARTOD-based automated QC tests
- Assist with program-wide cleanup and reorganization of OOI Raw Data Server (public-facing)
- Provide guidance to Keryx Project (Axiom redesign of UI) and OOI parent website
- Assessment of Roundabout (RDB) value for RCA asset management
- OOI Data Management Working Group and subgroups: QARTOD, Redmine, Communications
- Monitoring data flow from instruments through network to OOINet and Raw Servers
- Active investigation/resolution of data-related Redmine Tickets, including HelpDesk

# RCA PY2 QARTOD and HITL Workflow?

➔ Automated QC tests: \*\*\* usefulness highly dependent on reasonable and properly applied test parameters \*\*\*

## ► QARTOD:

- Gap test
- Syntax test
- Location test
- Gross range
- Climatology
- Spike test
- Rate of change test
- Flat line test
- Multi-variate test
- Attenuated signal test
- Neighbor test
- TS Curve/Space Test
- Density Inversion Test

QARTOD Manual											
In-situ Temperature and Salinity	Gap (Req)	Syntax (Req)	Location (Req)	Gross Range (Req)	Climatology (Req)	Spike (Strong)	Rate of Change (Strong)	Flat Line (Strong)	Multivariate (Sug)	Attenuated Signal (Sug)	Neighbor (Sug)
Passive Acoustics	Gap (Req)	Syntax (Req)	Location (Req)	Gross Range (Req)	Climatology (Req)	Spike (Sug)	Rate of Change (Sug)	Flat Line (Strong)	Multivariate (Sug)	Attenuated Signal (Sug)	Neighbor (Sug)
Phytoplankton	Gap (Req)	Syntax (Req)	Location (Req)	Gross Range (Req)	Climatology (Strong)	Spike (Strong)	Rate of Change (Strong)	Flat Line (Strong)	Multivariate (Sug)	Attenuated Signal (Sug)	Neighbor (Sug)
Dissolved Nutrients	Gap (Req)	Syntax (Req)	Location (Req)	Gross Range (Req)	Climatology (Strong)	Spike (Strong)	Rate of Change (Strong)	Flat Line (Strong)	Multivariate (Sug)	Attenuated Signal (Sug)	Neighbor (Sug)
Wind	Gap (Req)	Syntax (Req)	Location (Req)	Gross Range (Req)	Climatology (Req)	Spike (Strong)	Rate of Change (Strong)	Flat Line (Strong)	Multivariate (Sug)	Attenuated Signal (Sug)	Neighbor (Sug)
Water Level	Gap (Req)	Syntax (Req)	Location (Req)	Gross Range (Req)	Climatology (Req)	Spike (Strong)	Rate of Change (Strong)	Flat Line (Strong)	Multivariate (Sug)	Attenuated Signal (Sug)	Neighbor / Forecast (Sug)
Ocean Optics	Gap (Req)	Syntax (Req)	Location (Req)	Gross Range (Req)	Climatology (Strong)	Spike (Strong)	Rate of Change (Strong)	Flat Line (Strong)	Multivariate (Sug)	Attenuated Signal (Sug)	Neighbor (Sug)
Dissolved Oxygen	Gap (Req)	Syntax (Req)	Location (Req)	Gross Range (Req.)	Climatology (Req.)	Spike (Strong)	Rate of Change (Strong)	Flat Line (Strong)	Multivariate (Sug)	Attenuated Signal (Sug)	Neighbor (Sug)
In-situ Current	Gap (Sug)	Syntax (Req). This is similar to the checksum test	Location (Sug)	Gross Range (Sug). Could be applied to u and v velocities	Climatology (Sug). Could be applied to u and v velocities	Spike (Strong) applied to u and v velocities	Rate of Change (Strong) applied to u and v velocities	Flat Line (Req.)	Multivariate (Sug). Not included in manual	Attenuated Signal (Sug). Not included in manual	Neighbor (Sug). Not included in manual
In-situ Surface Water	LT Time Series Flat Line (Req)	LT Time Series Operational Freq (Req)	LT Time Series Low-Freq Energy (Req)	LT Time Series Bulk Wave Range (Req)	LT Time Series Rate of Change (Req)	Signal Strength (Strong)	Correlation Magnitude (Strong)	Acoustic Noise (Strong)	Signal to Noise (Strong)	Surface Tracking (Strong)	Velocity Min/Max (strong)

## Human in the Loop:

- ➔ Post-deployment checks to identify offsets
- ➔ Discrete sample verification
- ➔ Annotations

# QUESTIONS?