OCEAN OBSERVATORIES INITIATIVE

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

Wendi Ruef, Mike Vardaro, Orest Kawka, Rachel Scott, Kellen Rosburg

OOIFB Fall Meeting
Oct 13, 2020

RCA Data QA/QC: Completed Activities, PY2 Q3/4

- Participation in RCA O&M 2020 cruise:

  ▸ Collection and submission of discrete water samples for verification

  ▸ Initial verification of deployment and asset metadata

  ▸ Provided shore-based remote logging (Sealog)

- Completion of data-affecting Critical Metadata Audit on OOINet

- Investigated and resolved video feed and light issues for Axial HD Video Camera

- Initiated multiple sensor deep dives to investigate data issues (VEL3D/ADCP, DOSTAD)

- Investigated, tested, and resolved multiple data portal/backend issues (data catalog, raw archive)

- Initiated comprehensive historic annotation review and entry of new annotations

- Assessment of narrow and broad pCO2 calibration ranges on data

OOI OCEAN OBSERVATORIES INITIATIVE

NSF

RCA Data Team Outreach: PY2 Q3/4

- Presented at Ocean Sonics Live Stream Hydrophone Array Week (held remotely, 10/8/2020)

- Attended R2R advisory committee meeting (held remotely, 7/20/2020)

- Contributed pCO_2 data quality improvements article for OOI newsletter and main website

- Compiled RCA feedback and requirements for PMO initiatives: Keryx, Trapeze, RDB

- Compiled web data story on Axial Volcano Eruption Forecasting for new data portal

- Active investigation to resolve Help-desk user inquiries and data issues

OOI OCEAN OBSERVATORIES INITIATIVE

RCA Data QA/QC: Data Availability

- Timing Test:

  ▸ Auto-script to populate internal instrument status page and initiate email alerts

  ▸ M2M query returns most recent data timestamp ('endTime')

  ▸ Difference between current UTC time and 'endTime' determines criticality of "gap", dependent on instrument-specific thresholds


Fri, 09 Oct 2020 22:39:41 GMT

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
CE02SHBP-LJ01D-07-VEL3DC108		2020-10-09 22:31:42 UTC	0.0 sec
RS01SBPS-PC01A-05-ADCPTD102		2020-10-09 22:30:01 UTC	5.0 sec
RS01SBPS-PC01A-06-VADCPA101		2020-10-09 22:31:22 UTC	11.4 sec
RS01SBPS-PC01A-4A-CTDPFA103		2020-10-09 22:30:19 UTC	5.5 sec
RS01SBPS-PC01A-4A-DOSTAD103		2020-10-09 22:30:34 UTC	3.9 sec
RS01SBPS-PC01A-4B-PHSENA102		2020-10-09 22:00:00 UTC	31.3 min
RS01SBPS-PC01A-4C-FLORDD103		2020-10-09 22:30:43 UTC	1.8 sec
RS01SBPS-SF01A-2A-CTDPFA102		2020-10-09 22:30:22 UTC	1.0 sec
RS01SBPS-SF01A-2D-PHSENA101		2020-10-09 22:27:36 UTC	3.6 min
RS01SBPS-SF01A-3A-FLORTD101		2020-10-09 22:30:44 UTC	3.5 sec
RS01SBPS-SF01A-3B-OPTAAD101		2020-10-09 21:45:06 UTC	45.9 min
RS01SBPS-SF01A-3C-PARADA101		2020-10-09 22:31:01 UTC	4.0 sec
RS01SBPS-SF01A-3D-SPKIRA101		2020-10-09 22:31:26 UTC	1.7 sec
RS01SBPS-SF01A-4A-NUTNRA101		2020-10-09 21:49:33 UTC	41.3 min
RS01SBPS-SF01A-4B-VELPTD102		2020-10-09 22:31:39 UTC	3.8 sec
RS01SBPS-SF01A-4F-PCO2WA101		2020-10-09 22:24:46 UTC	6.4 min
RS01SLBS-LJ01A-05-HPIESA101		2020-10-09 22:22:57 UTC	7.9 min
RS01SLBS-LJ01A-10-ADCPTE101		2020-10-09 22:30:04 UTC	4.9 sec
RS01SLBS-LJ01A-11-OPTAAC103		2020-10-09 22:03:07 UTC	27.8 min
RS01SLBS-LJ01A-12-CTDPFB101		2020-10-09 22:30:30 UTC	0.7 sec
RS01SLBS-LJ01A-12-DOSTAD101		2020-10-09 22:30:35 UTC	2.1 sec
RS01SLBS-MJ01A-06-PRESTA101		2020-10-09 22:31:20 UTC	4.2 sec
RS01SLBS-MJ01A-12-VEL3DB101		2020-10-09 22:31:31 UTC	5.2 sec
RS03ASHS-MJ03B-07-TMPSFA301		2020-10-09 22:31:25 UTC	5.5 sec
RS03ASHS-MJ03B-09-BOTPTA304		2020-10-09 22:30:15 UTC	5.4 sec

OCEAN OBSERVATORIES INITIATIVE

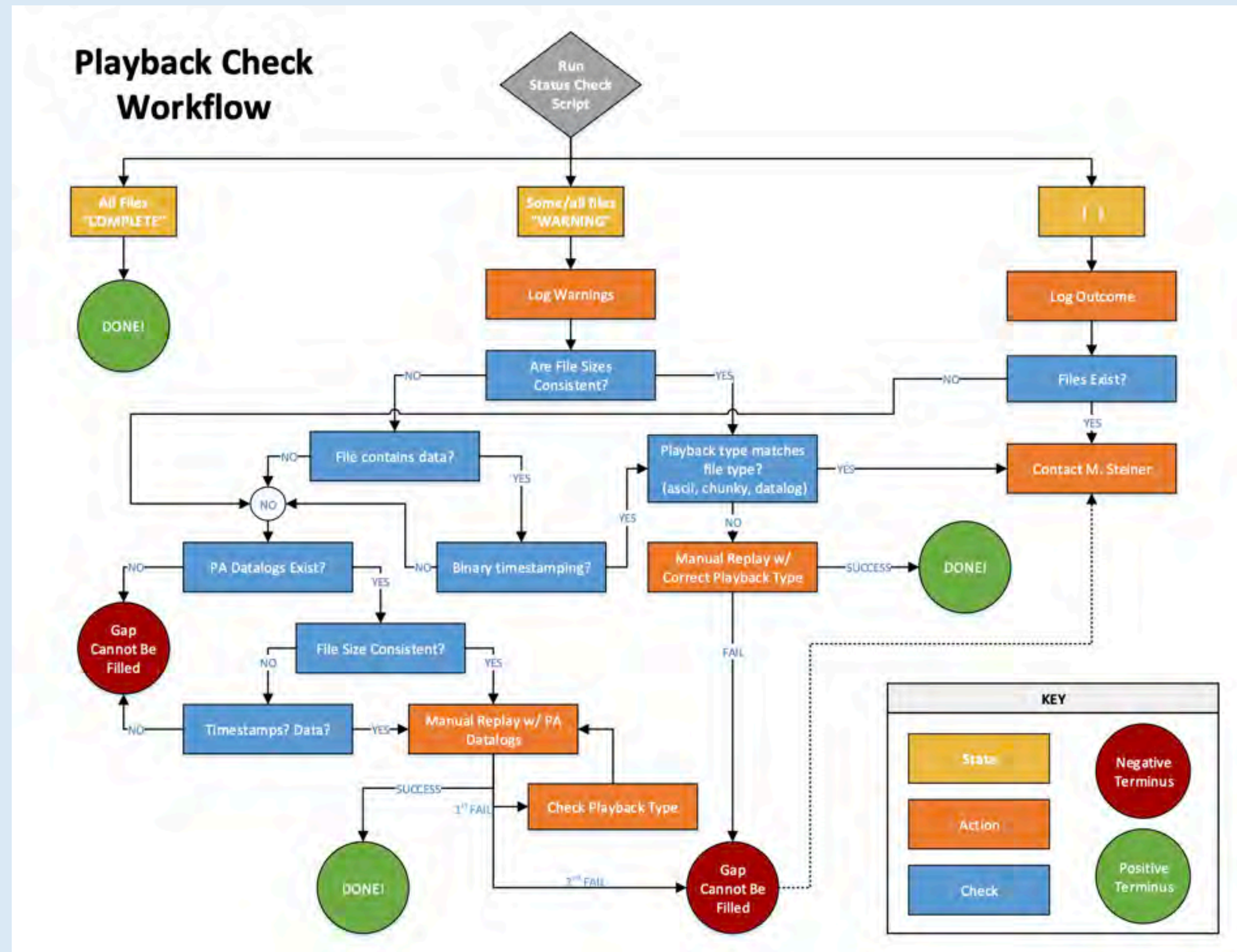# RCA Data QA/QC: Data Availability

- Timing Test:

  ▸ Auto-script to populate internal instrument status page and initiate email alerts

  ▸ M2M query returns most recent data timestamp ('endTime')

  ▸ Difference between current UTC time and 'endTime' determines criticality of "gap", dependent on instrument-specific thresholds
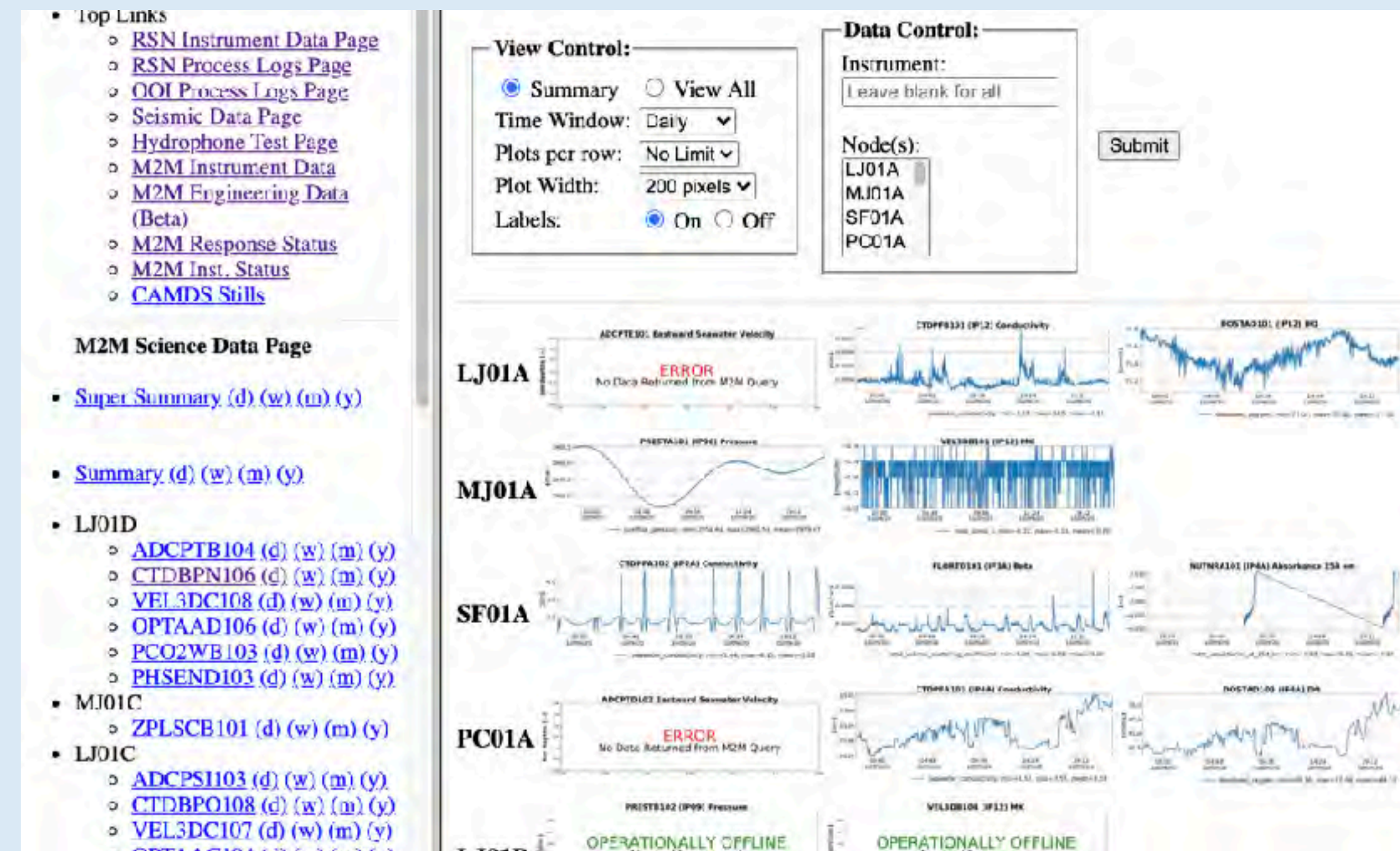
RCA Data QA/QC: Data Availability

- OOINet Data Gap Analysis:
  - ‣ Analyzes OOI data availability via M2M for individual reference designators by month
  - ‣ Detects gaps > 24 hours using a simple time gradient
  - ‣ Script can initiate playback for detected gaps using JSON format gap log
  - ‣ Expansion to gaps < 24 hours in progress, additional parameters, auto-annotatation

RCA Data QA/QC: Data Availability

- Raw Data Gap Analysis:

  ‣ Ultimate goal is to backfill gaps on the OOINet public-facing raw data server and complete playback to Cassandra; gap tool for UW raw archive but not for port agent logs

  ‣ Currently no comprehensive way to compare CI raw data archive, UW raw data archive, port agent logs, and Cassandra

  ‣ Hoping to work on solutions with CI over the coming months

RS01SBPS-PC01A-4A-CTDPFA103 2020-08-01 - 2020-09-01
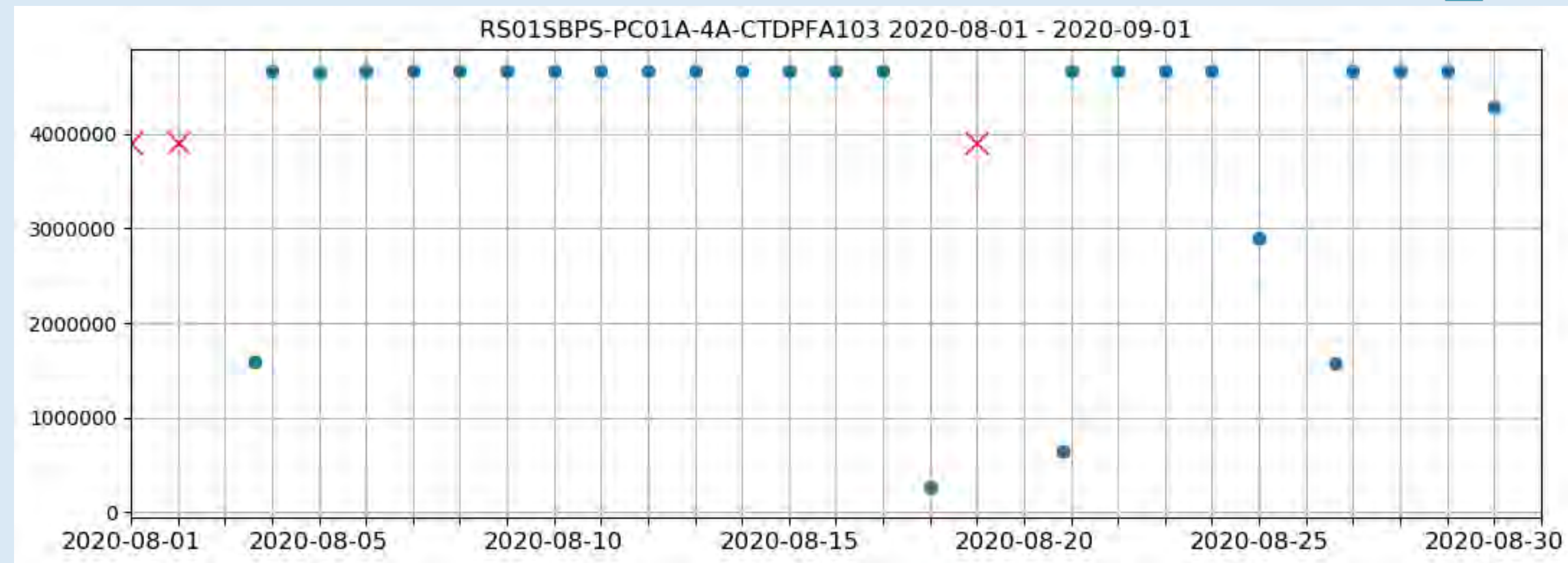
RCA Data QA/QC: Ongoing Activities

- Sensor-specific Deep Dives:

  ▸ Seafloor sensors:
    - Updates of instrument drivers and parsers for THSPH
    - Compilation of corrected TMPSF calibration coefficients, probe coordinates, headings

  ▸ Water-column sensor: Broadband Hydrophone (HYDBB)
    - Creation of transfer functions and calibration data repository
    - Resolution of HYDBB file fragmentation on raw archive
    - Implementation of additional data formats, capture and playback of diverted raw data
    - Increase of sampling rate from 64kHz to 256 kHz to meet science requirements
    - Visualization and access to audio data for hydrophones

  ▸ Bioacoustic Sonar (ZPLSC) visualization

  ▸ Test and implement Aanderaa oxygen optode cal phase calculation updates

OCEAN OBSERVATORIES INITIATIVE

RCA Data QA/QC: Ongoing Activities: IRIS Metadata

- Critical Metadata Audit of IRIS-based Tier One instruments

  ▸ OOI Instruments available through IRIS:
    - Low Frequency Hydrophones (HYDFLA): 5
    - Broadband Ocean Bottom Seismometers (OBSSBA): 5
    - Short-period Ocean Bottom Seismometers (OBSSPA): 9

  ▸ Metadata:
    - OBSSBA/HYDFLA: 17 channels
    - OBSSPA: 12 channels

  ▸ Total coefficients to verify > 200

OCEAN OBSERVATORIES INITIATIVE

RCA Data QA/QC: Ongoing Activities: IRIS Metadata

- Workflow:

  ▸ Sensor coefficients provided by vendor for data loggers and sensors

DM24 CALIBRATION

WORKS ORDER: 13071   DIGITISER SERIAL NUMBER: A4335

SYSTEM ID: 13071	CPLD: A0.E1
UNIT ID: 4335	BOOTLOADER: MK3BOOT213.IMG
OUTPUT DATA FORMAT: GCF	DSP SOFTWARE: DSP1090.BIN
BAUD RATE: 38400	SYSTEM: DMNET106b57d.IMG

VELOCITY CHANNELS

Channel:	4335Z2	Vertical	2.873 µV/Count
	4335N2	North/South	2.874 µV/Count
	4335E2	East/West	2.880 µV/Count

MASS POSITION CHANNELS

Sample Rate: 4 samples/sec (Default)

Channel:	4335M8	Vertical	290.317 µV/Count
	4335M9	North/South	290.466 µV/Count
	4335MA	East/West	290.787 µV/Count

Sample Rate: 1 samples/sec

Channel:	4335M8	Vertical	2.268 µV/Count
	4335M9	North/South	2.269 µV/Count
	4335MA	East/West	2.272 µV/Count

OCEAN OBSERVATORIES INITIATIVE
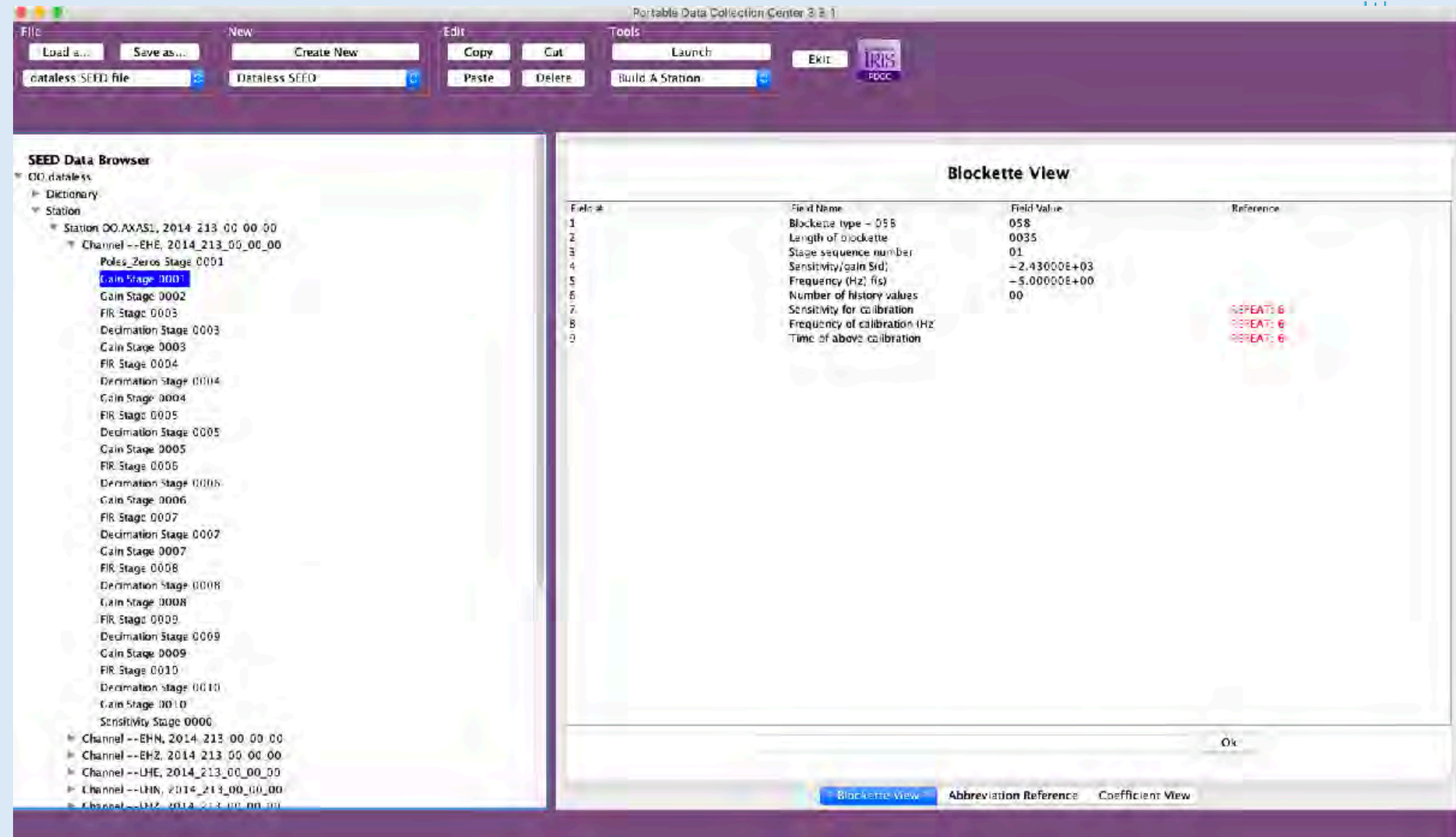
RCA Data QA/QC: Ongoing Activities: IRIS Metadata

- Workflow:

  ▸ Sensor coefficients provided by vendor for data loggers and sensors

  ▸ Each coefficient undergoes manual unit transformation with 2i-HITL verification

Asset ID		ATAPL-58331-00001	ATAPL-58331-00002	ATAPL-58331-00003	ATAPL-58331-00004
Serial Number		T6J67	T6J68	T6J69	T6J70
EAM		5163	5164	5165	5166
Digitiser		A4335	A4330	A4334	A4414
	Units	Calibration Coefficients			
Z2	uV/Count	2.873	2.873	2.873	2.871
	count/V	348068.2214	348068.2214	348068.2214	348310.6931
N2	uV/Count	2.874	2.873	2.872	2.871
	count/V	347947.112	348068.2214	348189.415	348310.6931
E2	uV/Count	2.88	2.872	2.873	2.871
	count/V	347222.2222	348189.415	348068.2214	348310.6931
M8: 4 Hz	uV/Count	290.317	290.943	291.862	290.099
	count/V	3444.510656	3437.099363	3426.276802	3447.099094
M9: 4 Hz	uV/Count	290.466	290.965	291.173	291.816
	count/V	3442.743729	3436.839482	3434.384369	3426.816898
MA: 4 Hz	uV/Count	290.787	290.601	290.98	291.363
	count/V	3438.943282	3441.144387	3436.662314	3432.144782

- Workflow:

  ‣ Sensor coefficients provided by vendor for data loggers and sensors

  ‣ Each coefficient undergoes manual unit transformation with 2i-HITL verification

  ‣ Verified coefficients are entered into IRIS PDCC dataless SEED file

OCEAN
OBSERVATORIES
INITIATIVE
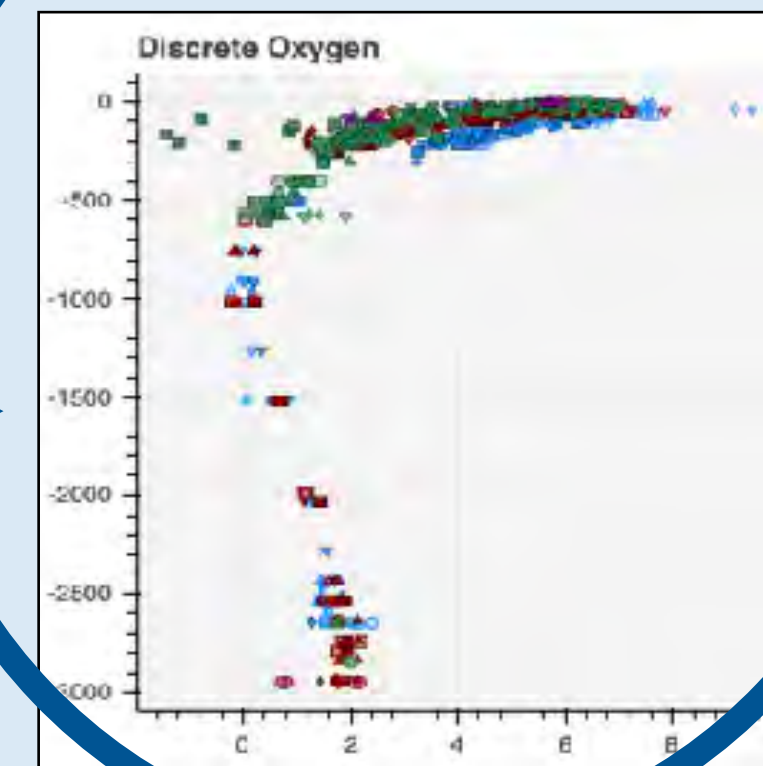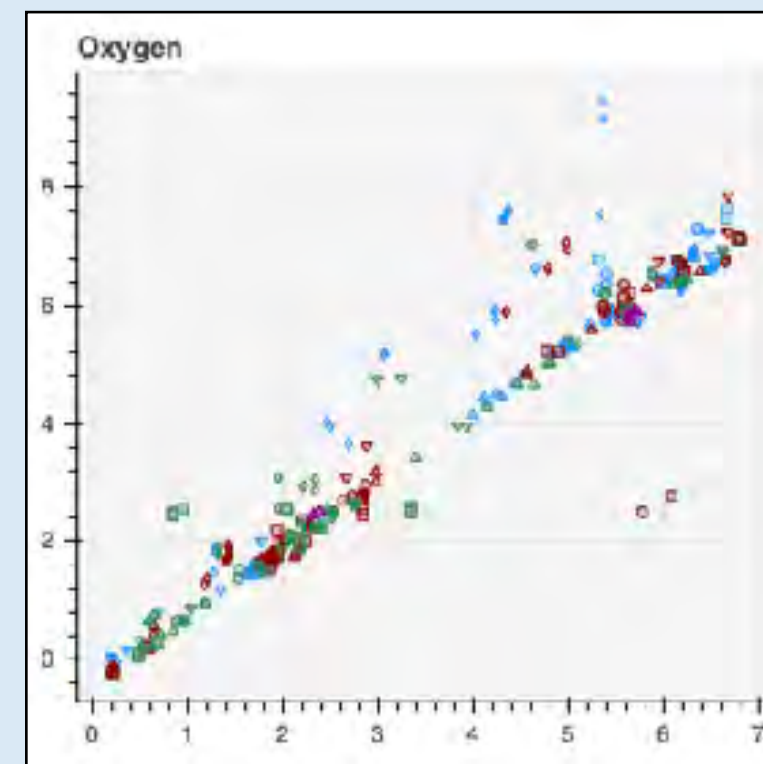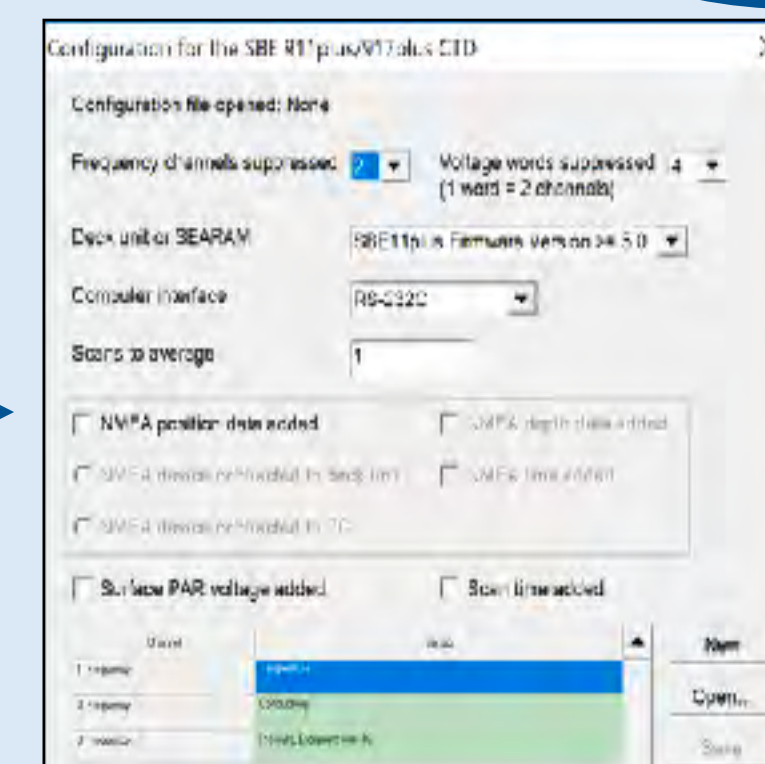
RCA Data QA/QC: Ongoing Activities: Discrete Samples
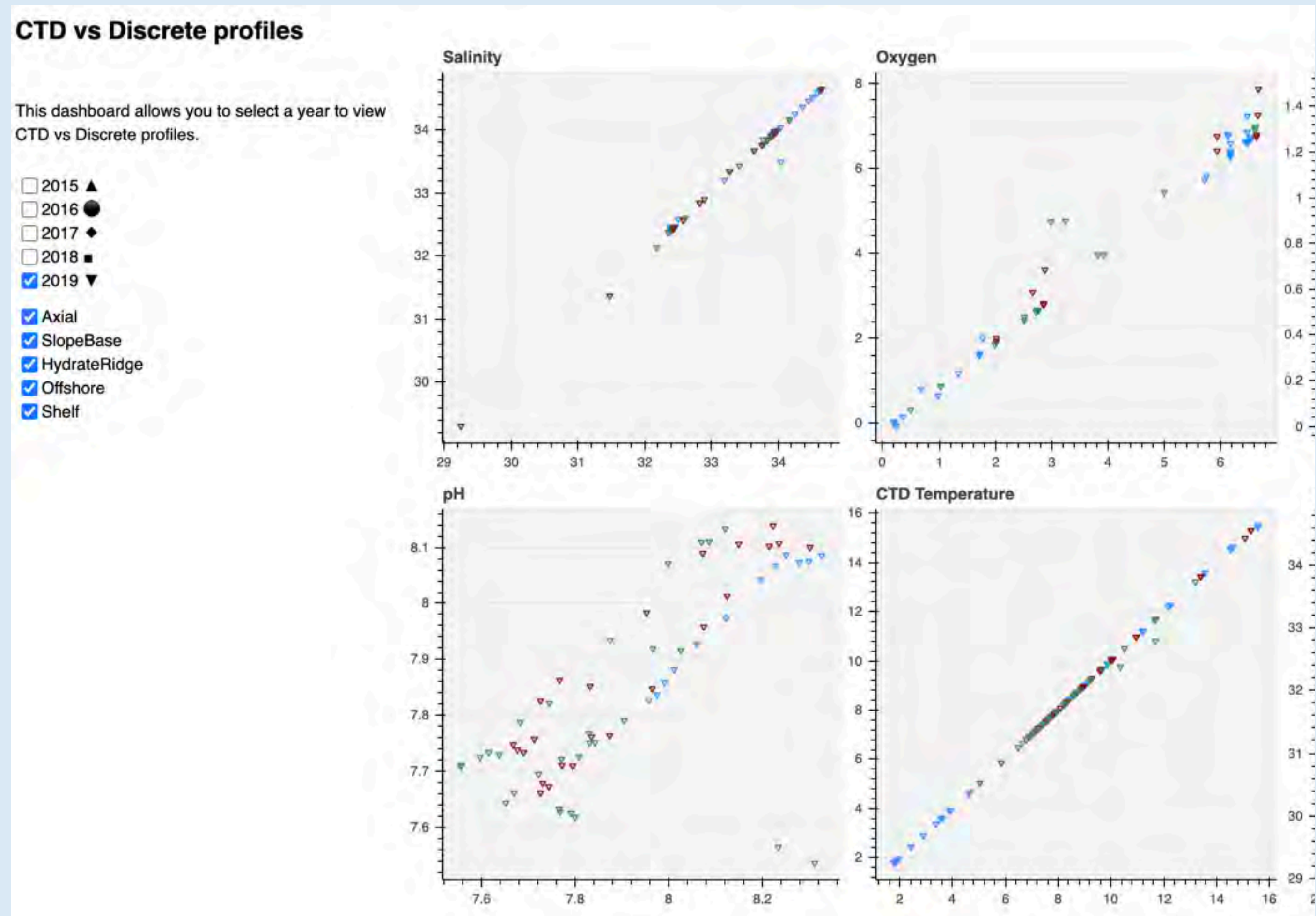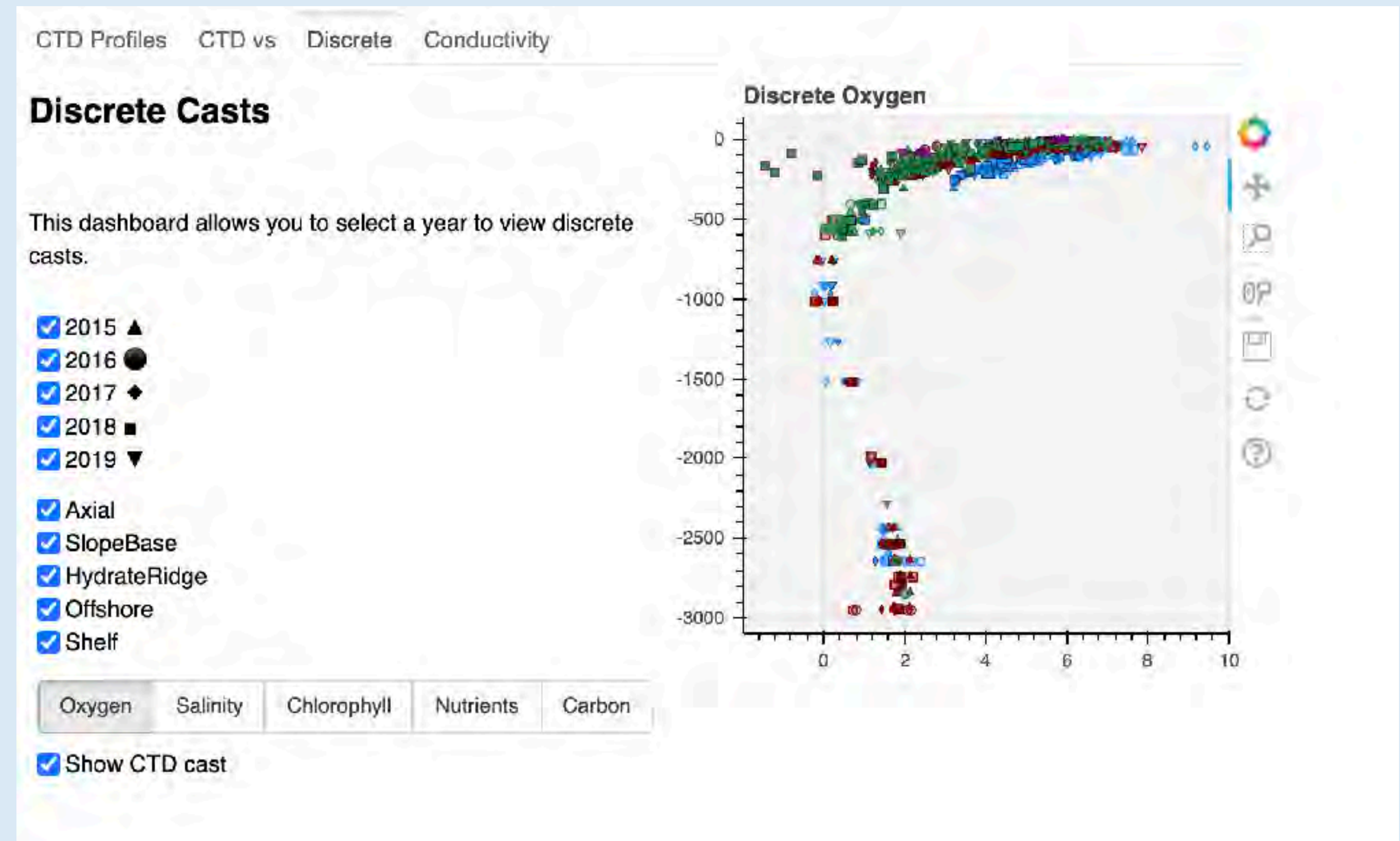
RCA Data QA/QC: Ongoing Activities: Discrete Samples

- Dashboard:

  ‣ Quick viewing and comparison of discrete samples, shipboard CTD profiles, and array data

  ‣ RCA array data harvested from UW cloud-based zarr files

  ‣ Could be adapted to a web-based user tool

OCEAN OBSERVATORIES INITIATIVE

- Dashboard:

  ▸ Quick viewing and comparison of discrete samples, shipboard CTD profiles, and array data

  ▸ RCA array data harvested from UW cloud-based zarr files

  ▸ Could be adapted to a web-based user tool

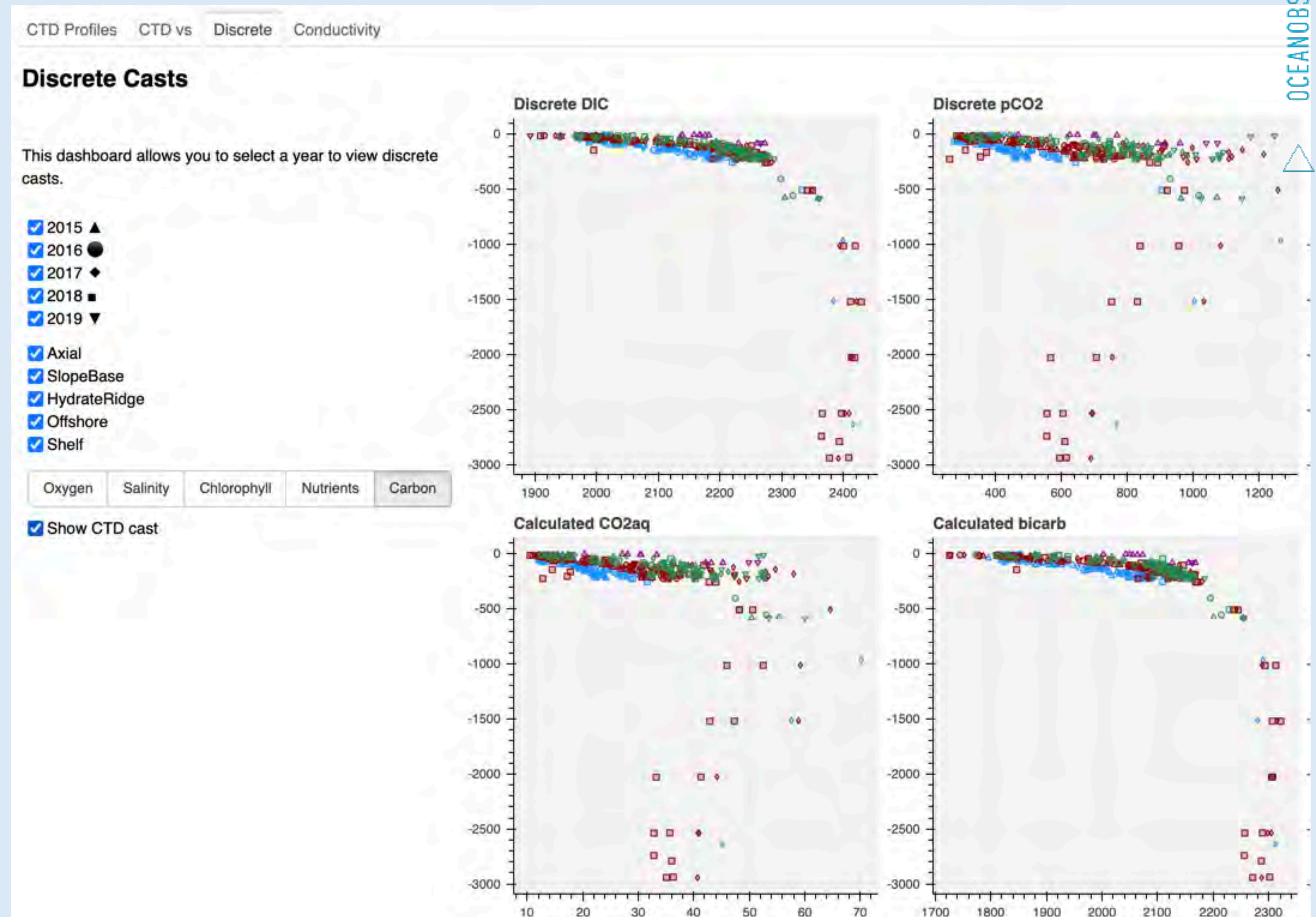
OCEANOBSERVATORIES.ORG

OCEAN OBSERVATORIES INITIATIVE

- Dashboard:

  ▸ Quick viewing and comparison of discrete samples, shipboard CTD profiles, and array data

  ▸ RCA array data harvested from UW cloud-based zarr files

  ▸ Could be adapted to a web-based user tool
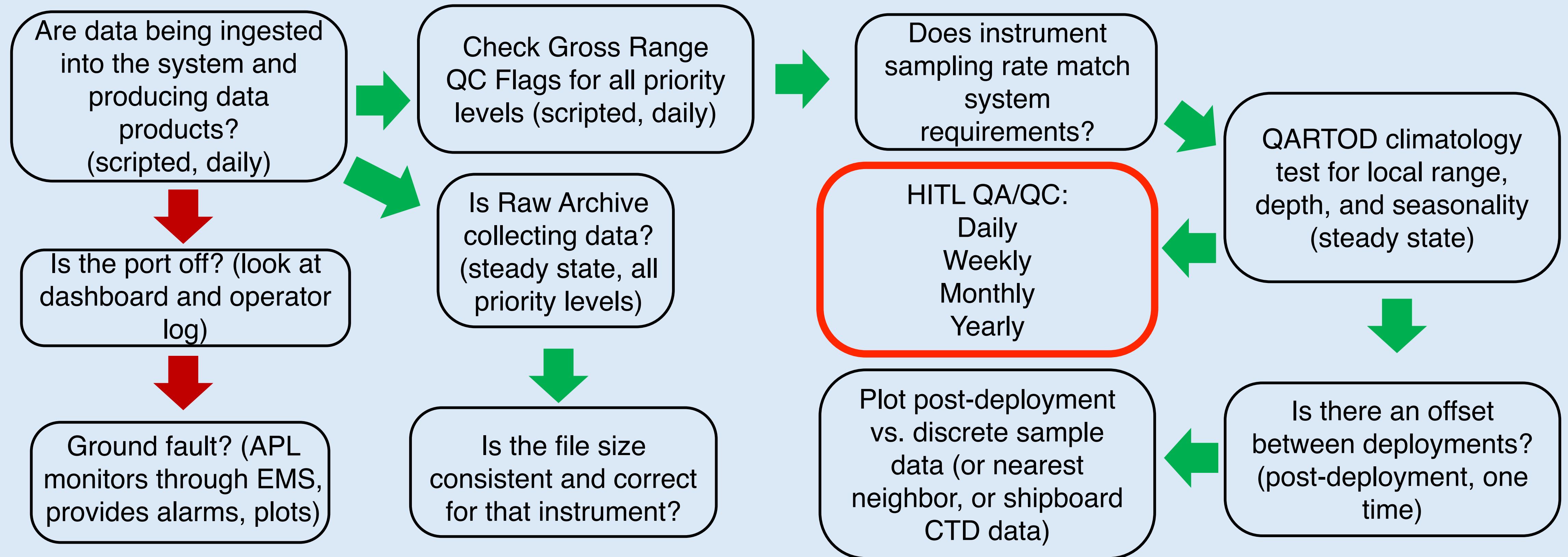

OCEAN OBSERVATORIES INITIATIVE

RCA Data QA/QC HITL Workflow (DRAFT)

Are data being ingested into the system and producing data products? (scripted, daily)

Check Gross Range QC Flags for all priority levels (scripted, daily)

Does instrument sampling rate match system requirements?

QARTOD climatology test for local range, depth, and seasonality (steady state)

Is the port off? (look at dashboard and operator log)

Is Raw Archive collecting data? (steady state, all priority levels)

HITL QA/QC:
Daily
Weekly
Monthly
Yearly

Ground fault? (APL monitors through EMS, provides alarms, plots)

Is the file size consistent and correct for that instrument?

Plot post-deployment vs. discrete sample data (or nearest neighbor, or shipboard CTD data)

Is there an offset between deployments? (post-deployment, one time)

- If an issue occurs more than once within a 24-hour period, it receives a single annotation
- If an issue occurs more than 24 hours after the first issue occurred, it receives a separate annotation
- If a time period has frequent issues, the entire week/month/deployment receives a single annotation

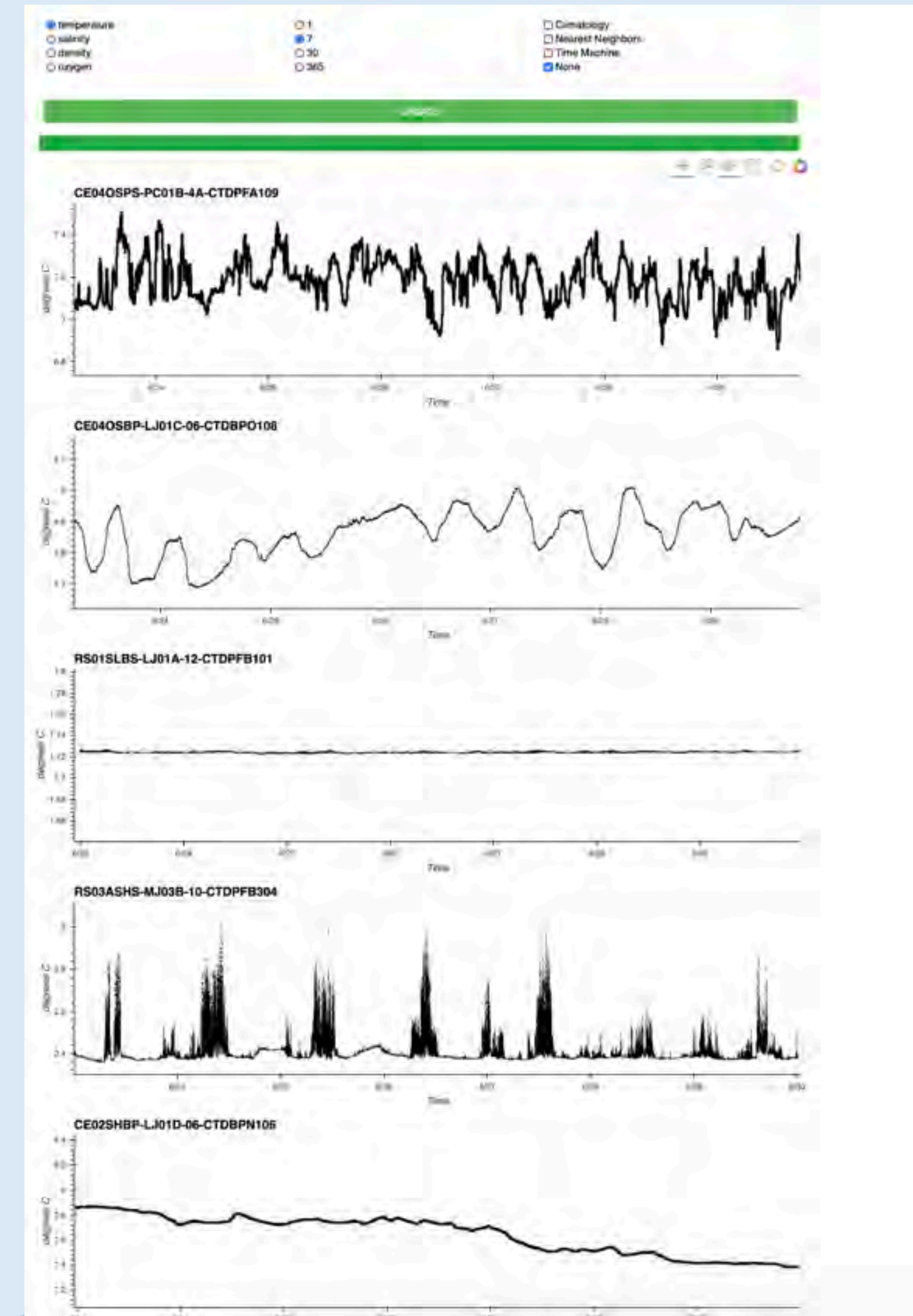# RCA Data QA/QC: Ongoing Activities: HITL QA/QC

- Instrument Categorization for QC checks based on user priority

  ▸ **Daily:** CTD, DO, BOTPT, NUTNR, FLOR, pCO2, pH, PAR, SPKIR, ADCP, VELPT

  ▸ **Weekly**: CAMHD, THSPH, TRHPH, TMPSF, ZPLSC, OPTAA, PREST, VEL3D, VADCP

  ▸ **Bi-Weekly**: HPIES, CAMDS, HYDLF, HYDBB, OBSBB, OBSSP

  ▸ **Annually**: Physical samples, Core Analytical Data

  ▸ **External SMEs:** PI instruments

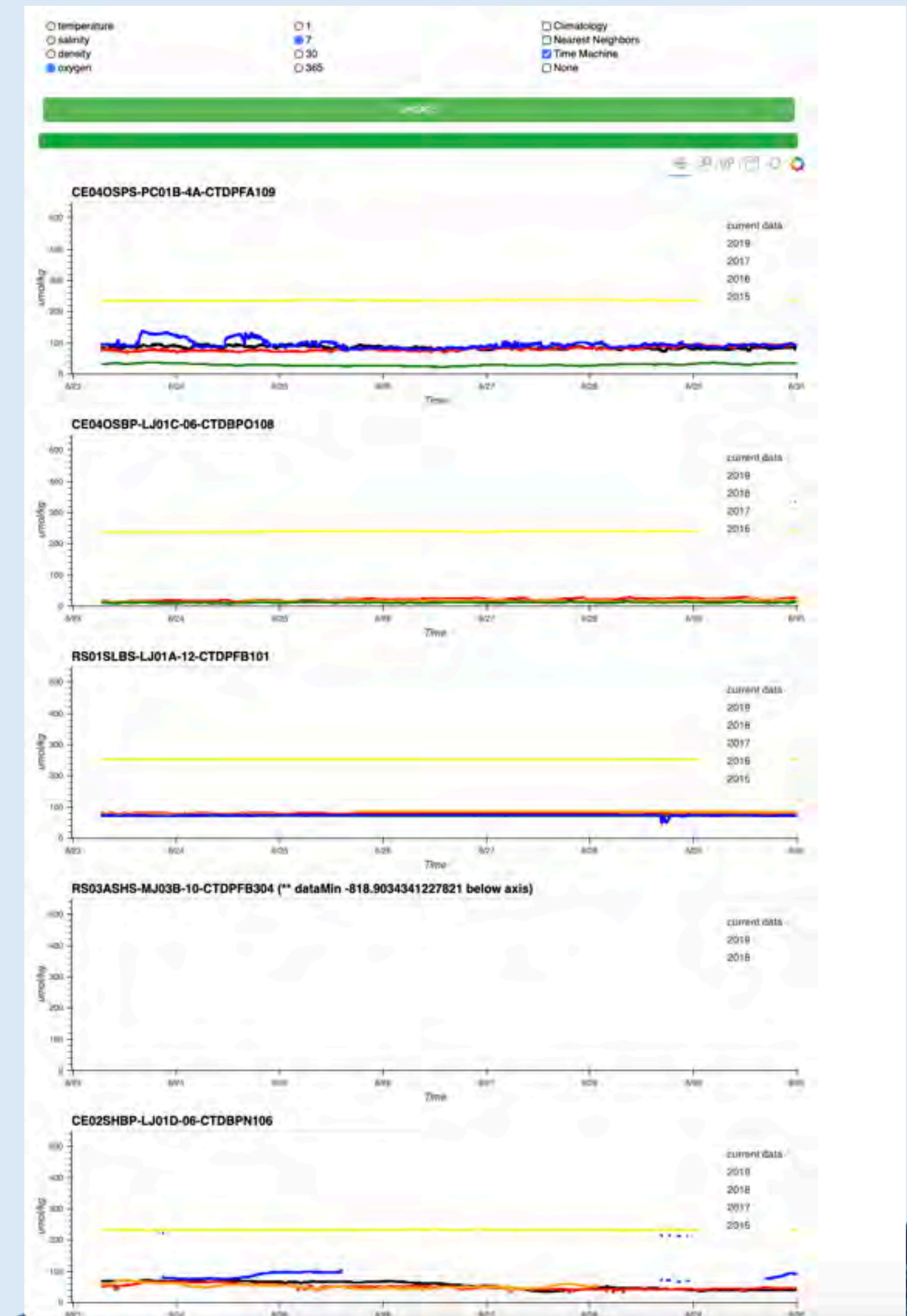# RCA Data QA/QC: Ongoing Activities: HITL QA/QC

- QA/QC Dashboard:

  ▸ Quick view of assets on one page, viewable by variable (i.e. temperature) or site (i.e. Axial)

  ▸ RCA array data harvested from UW cloud-based zarr files to allow for fast access and interactivity of multiple large datasets

  ▸ Overlay climatology, nearest neighbors, previous data, QC flags

  ▸ Rotate between set timespans (day, week, month, year)

- QA/QC Dashboard:

  ▸ Quick view of assets on one page, viewable by variable (i.e. temperature) or site (i.e. Axial)

  ▸ RCA array data harvested from UW cloud-based zarr files to allow for fast access and interactivity of multiple large datasets

  ▸ Overlay climatology, nearest neighbors, previous data, QC flags

  ▸ Rotate between set timespans (day, week, month, year)

OCEAN OBSERVATORIES INITIATIVE

OCEANOBSERVATORIES.ORG

RCA Data QA/QC: Ongoing Activities

- Assisting with design and implementation of QARTOD-based automated QC tests

- Cross-MIO collaboration to update OOI Documentation (e.g. DPS documents)

- Emergent Sensor Deep Dives

- Ongoing periodic review and maintenance of Critical Metadata

- OOI Data Management Working Group and subgroups: QARTOD, Redmine, Preload, Keryx, RDB, DPS Updates, Sensor-specific (ADCP, Pressure, Bioacoustic Sonar (ZPLSC) Data Products)

- Monitoring data flow from instruments through network to OOI and RCA Raw Archives

- Active investigation/resolution of data-related Redmine Tickets, including Help-desk

- Annotation review and entry of new annotations

- Posting current data issues and outages to oceanobservatories.org
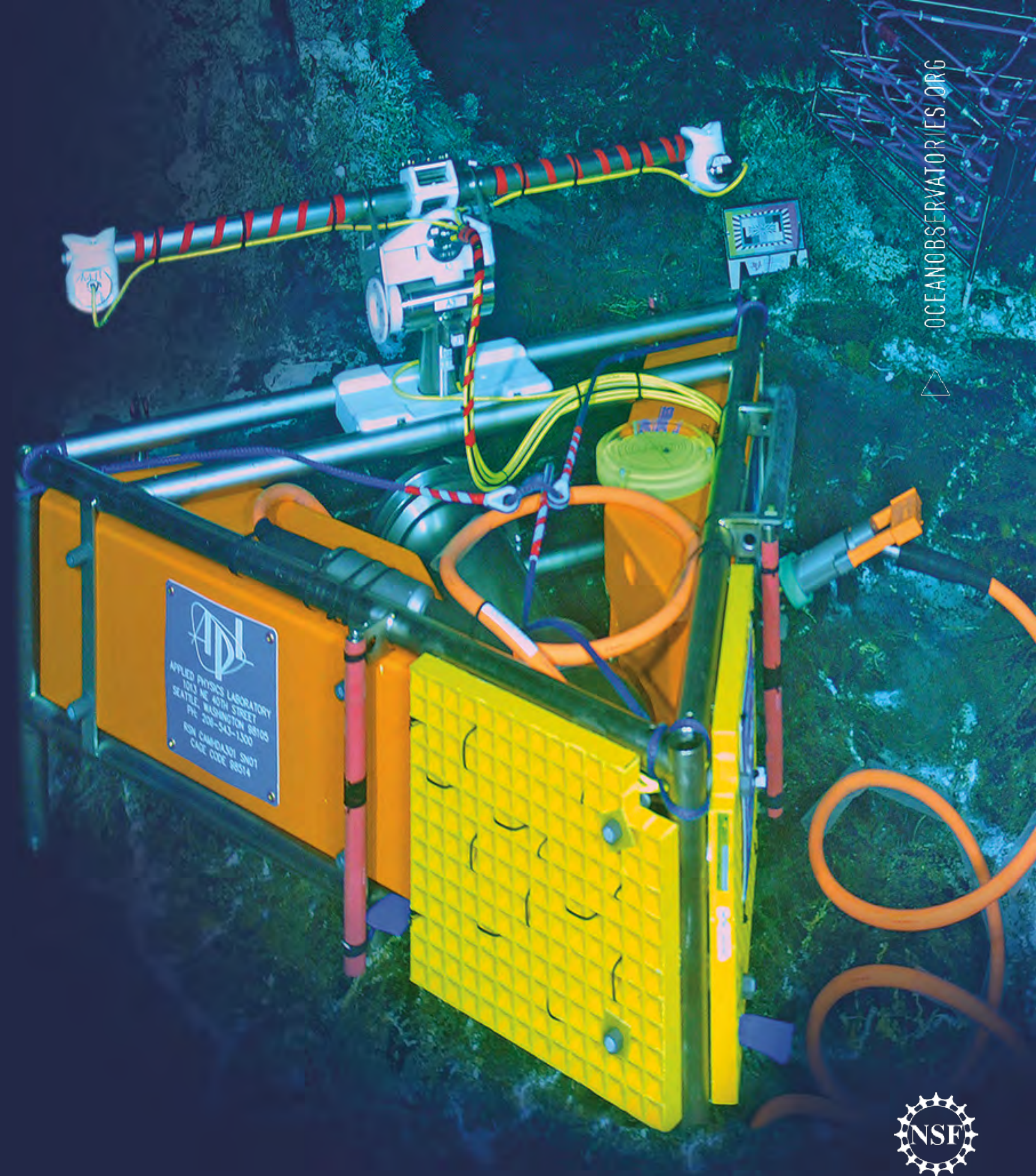
OCEAN
OBSERVATORIES
INITIATIVE

RCA Data QA/QC: Upcoming Priorities

- Completion of RCA historical data availability check on OOINet, expansion of capabilities

- Completion of RCA historical data availability check on UW Raw Data Archive

- Completion of Critical Metadata Audit of IRIS-based Tier One metadata

- Finalization and Implementation of RCA-specific HITL data QA/QC workflow

- Development of documentation for metadata verification status and HITL QA/QC for users

- Initial data verification of RCA 2020 instruments using ship-based CTD profiles/discrete samples

- Completion of historic annotation review

- Broadband Hydrophone deep dive and development

- Completion of final verification for RCA O&M 2020 cruise and critical metadata

- Compilation of Core Analytical Data and Metadata products

OOI OCEAN OBSERVATORIES INITIATIVE

Questions?