Marine Data Wrangling with OpenRVDAS and CORIOLIX
Part 2: CORIOLIX

- Introduction
  - What is it?
  - What does it do?
  - Why do we need it?
- How does it work?
  - Architecture / Workflow / Requirements
- How does the OpenRVDAS<-> CORIOLIX part work?
  - Managing Logging Configuration
  - Performing Real-Time Transformations
  - Public CORIOLIX Resources (Repositories):
    - Message Parsing Expressions
    - Parameters
    - Parameter Transformations
    - Flagging Routines

CORIOLIX: The OSU support team

Chris Romsos

Datapresence Systems Engineer
Lead for Datapresence and RCRV Cyberinfrastructure
Background – Marine Geology, Fisheries, Geographic Info Systems

Jasmine Nahorniak

Assistant Datapresence Systems Engineer
Lead CORIOLIX developer
Background – Ocean Optics, Data Processing & Management, UI design

Tristan King

Data Systems Specialist
Development + Manage CORIOLIX installations on vessels
Background – 3D Visualization and Web Development

CORIOLIX Installations

Currently:

- Point Sur
- Sikuliaq
- Savannah
- Roger Revelle (in progress)
- Okeanos Explorer (in progress)

Next Up (Fall/Winter 25/26):

- Bell Shimada
- Pelican

Retired:

- Oceanus
- Endeavor









What Is CORIOLIX?

OpenRVDAS: Part of suite of open source tools for data acquisition/management under the "Ocean Data Tools" collaboration.

CORIOLIX, In one line -

- A distributed software system, connecting shipboard and shoreside nodes, for live situational awareness.

Expanding on above, CORIOLIX -

- Is designed around real-time/near-real-time mission support: not just data logging but live visualization, alerts, remote monitoring.
- Bridges ship-to-shore in a unified way: same interface for both, which helps coordination. Moves real-time observational data to shore and near real-time products to ship (bi-directional)
- Built with operations in mind (sensor health monitoring, event logs, remote alerts) — adds operational value beyond just science data capture.

What Is CORIOLIX?

Jargon translator

A ***distributed software system***, connecting shipboard and shoreside ***nodes***, for ***live situational awareness***.

Translations:

Distributed software system: collection of independent computers, or nodes, working together over a network toward a common goal and appearing to the end user as a single, coherent system.

Nodes: independent computers, see above.

Live: real-time (or near real-time).

Situational Awareness: Describes an ability to perceive, understand, and predict the environment and it's elements to make informed decisions.

Feature Comparison:

Category	OpenRVDAS	CORIOLIX
Main Function	Collects & logs sensor data	Displays, monitors & shares data ship/shore
Primary Users	Techs configuring sensors & systems	Techs, scientists, and shore observers
Architecture	Modular pipeline: Reader → Transform → Writer	Web app: dashboards, charts, maps, logs
Setup & Maintenance	Requires tech config; YAML/scripts	Mostly web-based; setup takes effort
Data Inputs	Serial, network, NMEA, binary sensors	Pulls from DB/API (e.g., OpenRVDAS output)
Data Outputs	Log files, InfluxDB, Grafana feeds	Dashboards, data services, event & sensor logs
Real-Time Visualization	Via Grafana (custom setup)	Built-in, ready-to-use interface for ship & shore
Ship-to-Shore	Add-on via sync tools	Native feature; shore access built-in
Event Logging	Limited (system logs)	Integrated cruise/event logbook

What Does It Do?

Data Acquisition:

- Manages your logging configuration and logging state across the full system.

Data Processing:

- Performs both real-time and delayed/batch processing (transformations, flagging, binning)

Data Storage & Management:

- Annotates data, logs system configuration, performs data replication between ship and shore

Data Access:

- Offers multiple modes (file download, real-time streaming, API polling)
- Provides varied output formats (ascii, JSON, OpenDAP, netCDF, OGC, MQTT)
- Data visualization (Timeseries maps, charts, and graphs)
- Supports role-based user access

Does it do all that on its own?

Heck No!

CORIOLIX integrates several open source software tools. A few of the major ones include:

- Data Acquisition: OpenRVDAS or SCS (w/limited functionality)
- Scientific Data Server: ERDDAP (query, subset, format translate, OGC services)
- Database: PostgreSQL w/PostGIS & Timescale extensions
- Database Replication: SymmetricDS
- Mapping: Leaflet
- Charting: Apache ECharts (new!, replaces Highcharts)

OpenRVDAS and CORIOLIX: Two Tools, One Workflow

- *OpenRVDAS:* Data acquisition, processing, and logging
- *CORIOLIX:* Visualization, monitoring, and ship-to-shore access

Together: *End-to-end data visibility from sensors to shore*

CORIOLIX Front End Tour

Public access:

- Visualize and Download Data
- View Sensor Metadata
- View Event and Sensor Logs

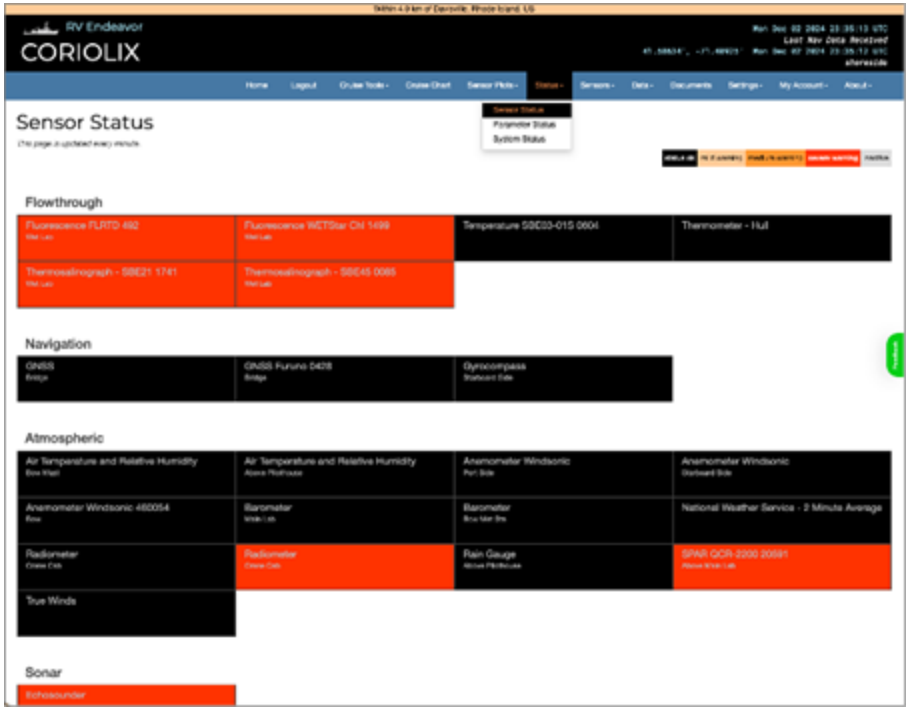
Scientist account access:

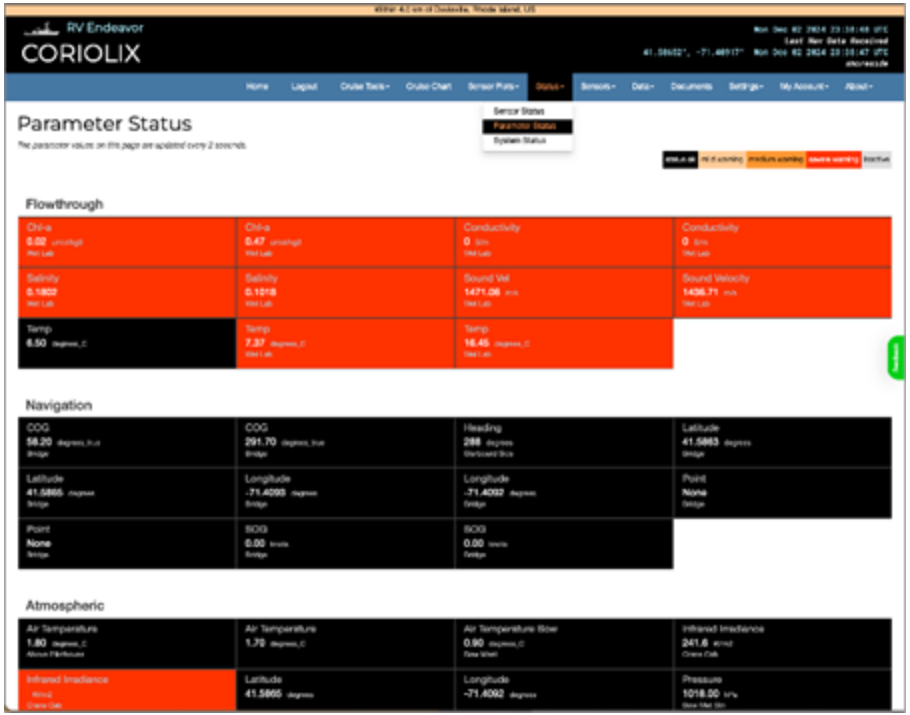
- Visualize and Download Data
- View Sensor Metadata
- View Sensor Logs
- Create Event Logs
- Set Personal Notifications

Technician account access:

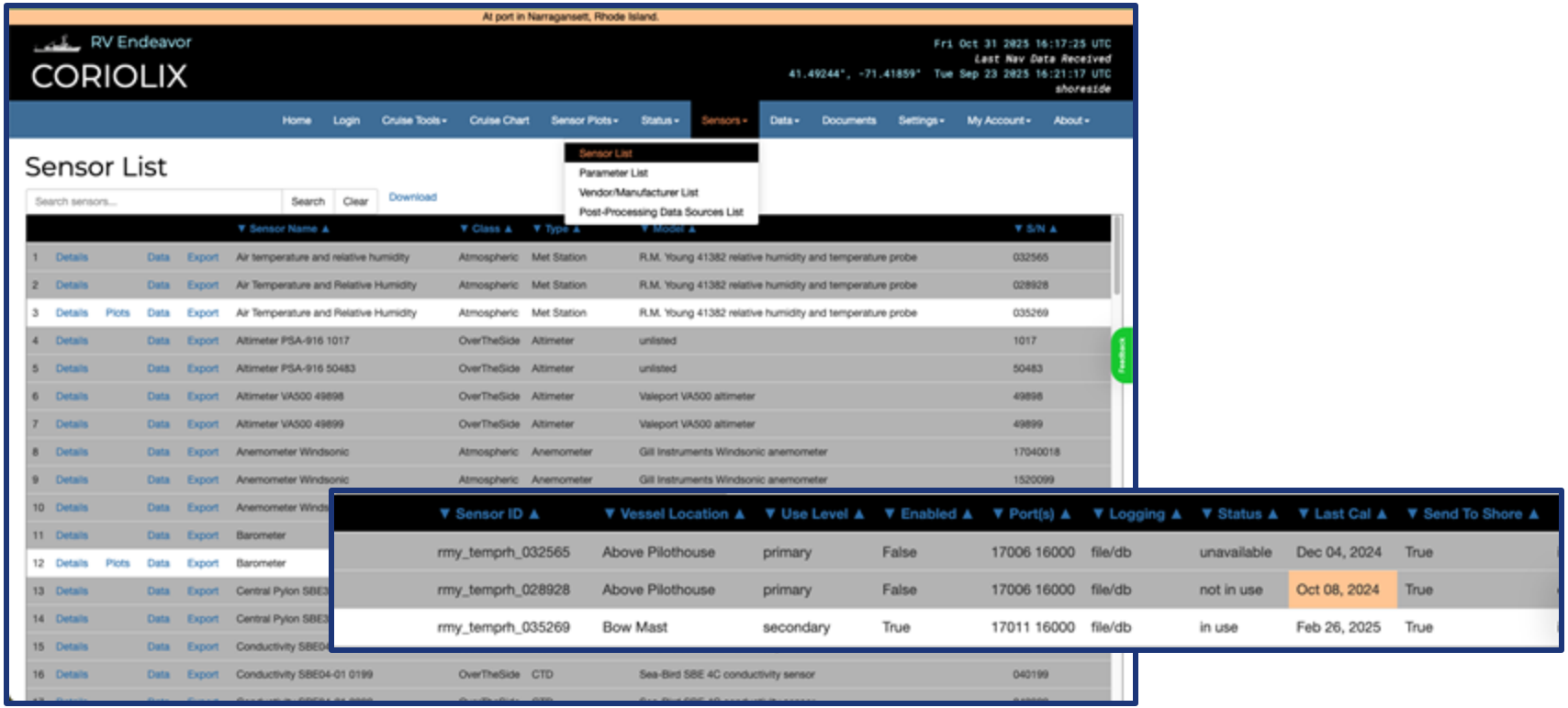
- Visualize and Download Data
- Create/Edit Sensor Metadata
- Create/Edit Event and Sensor Logs
- Change System Configuration
- Set Personal Notifications

What's the status of my system?

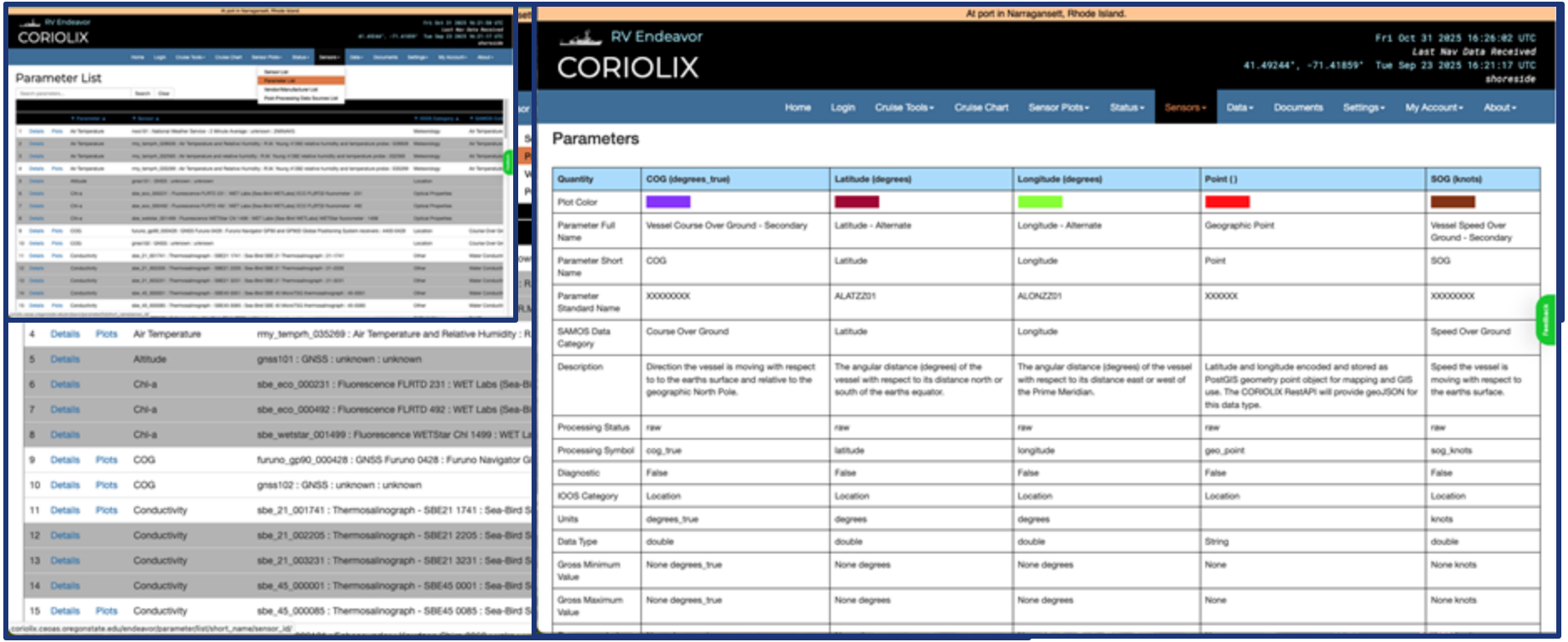




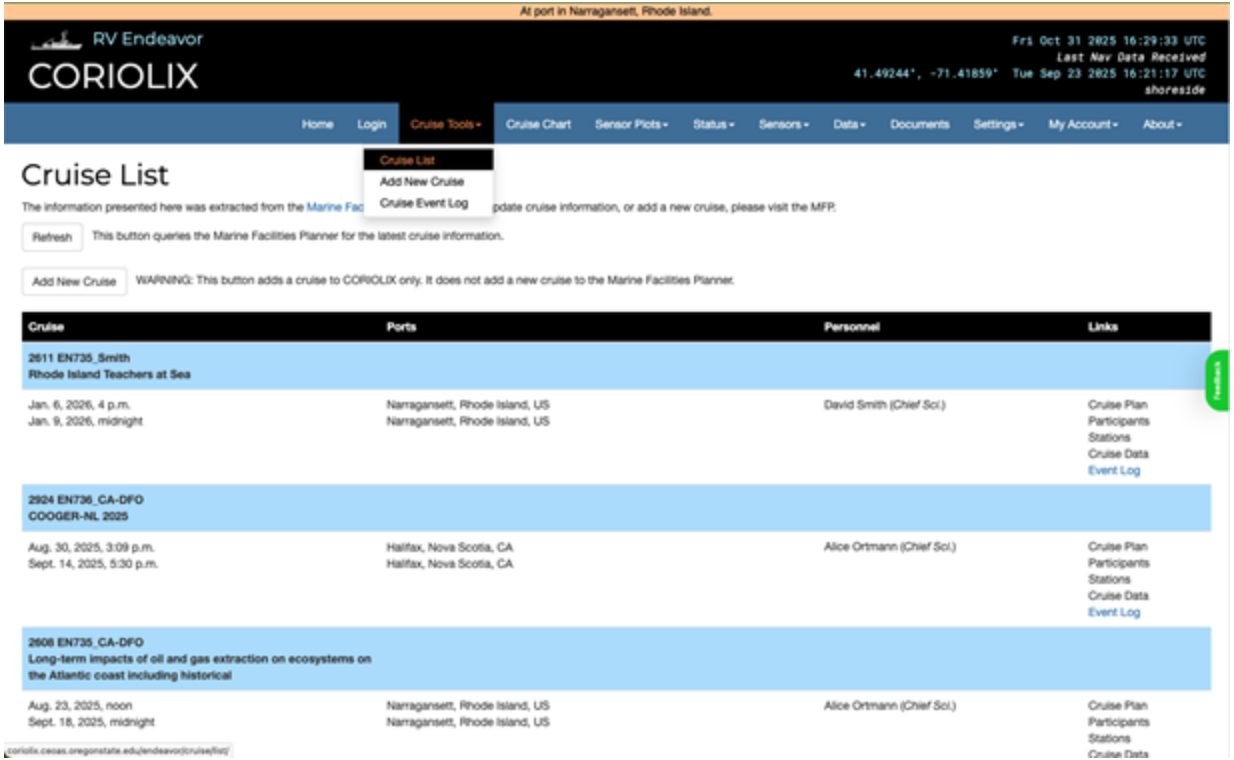
Sensors Inventory



Parameters measured

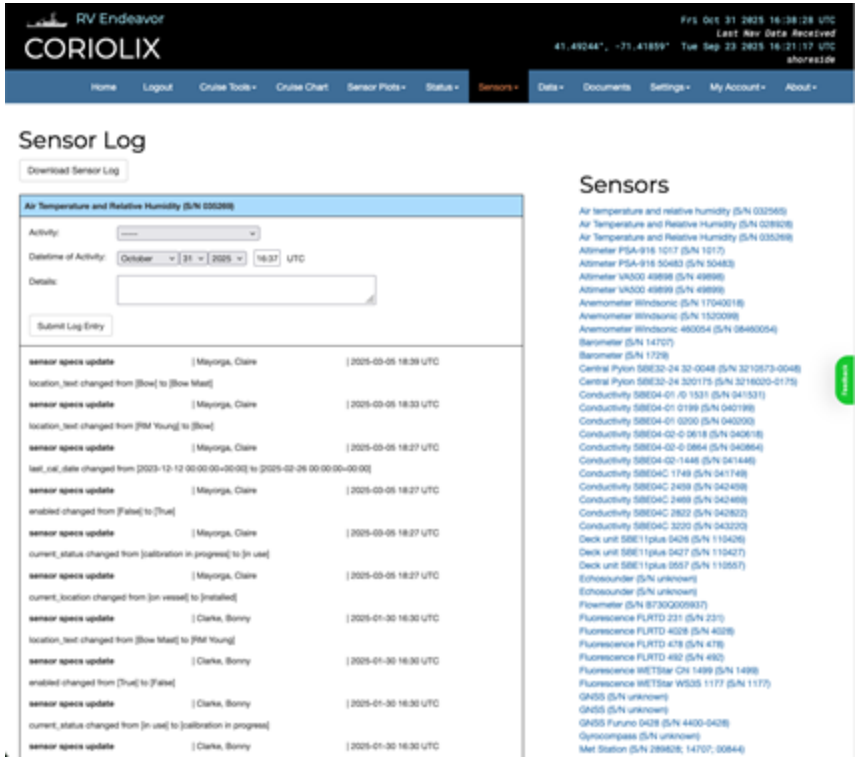


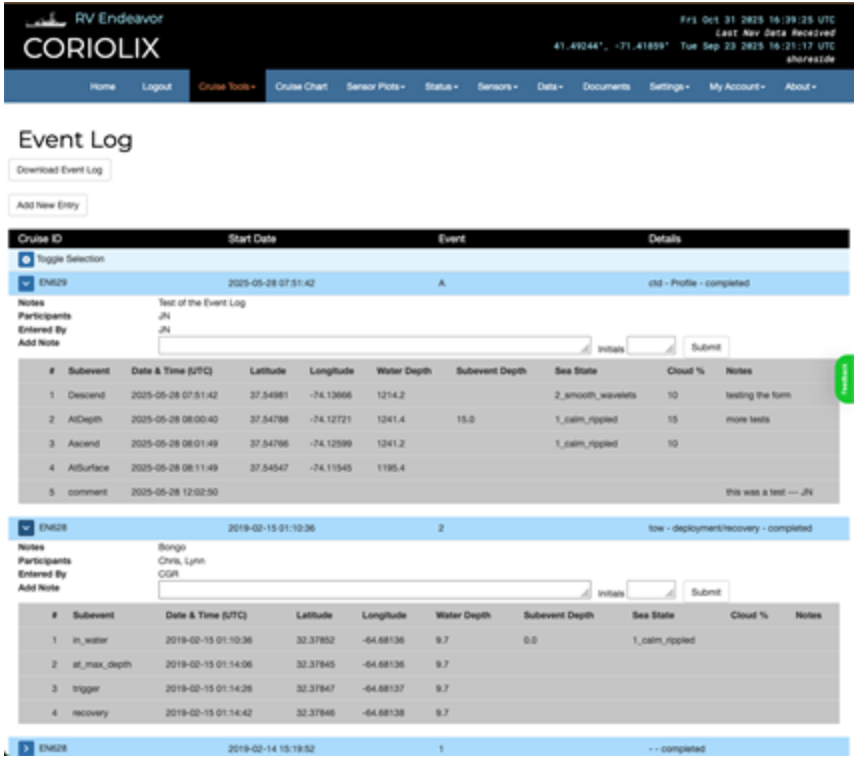
MFP & MOMS: Pulls from source of truth



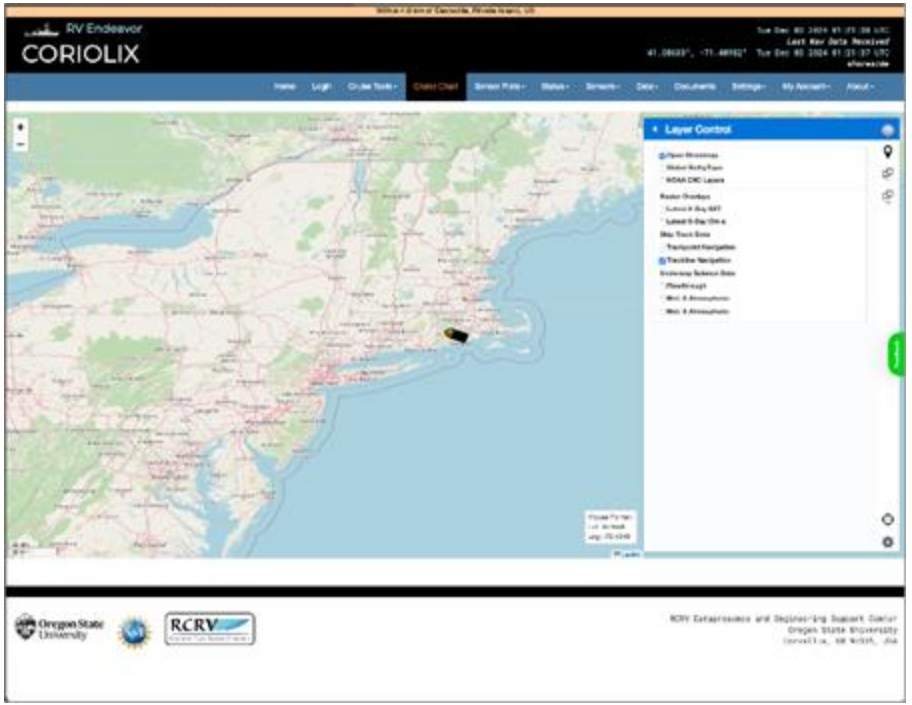
planning metadata

Sensor & Event Logs.





Data Visualization



System Requirements

Assumes you can obtain, create, or have:

- SatComs
- Signed Server Certificates
- DNS Records
- Firewall rules to allow traffic to flow

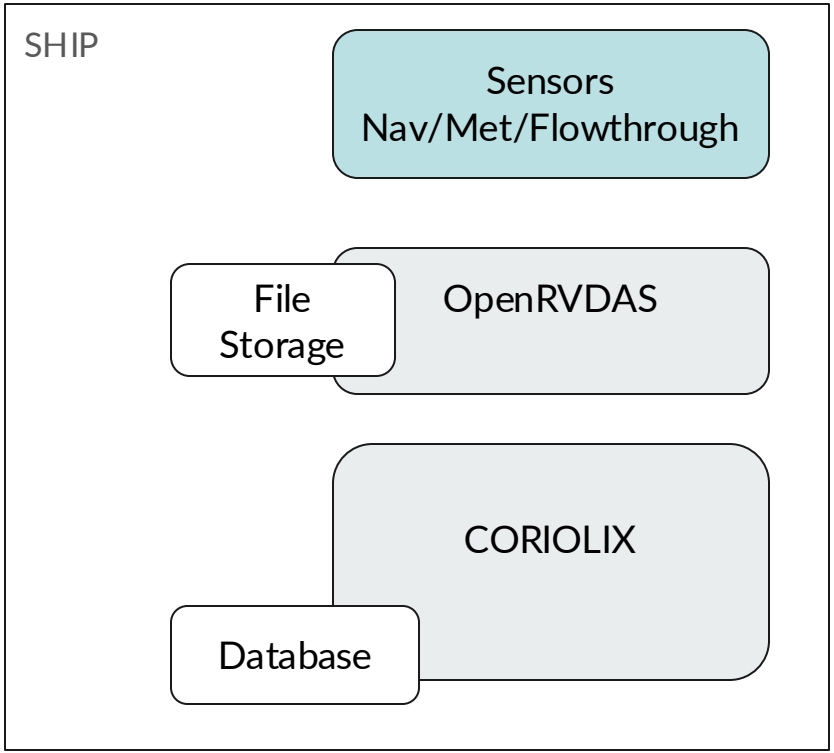
Site & Hardware:

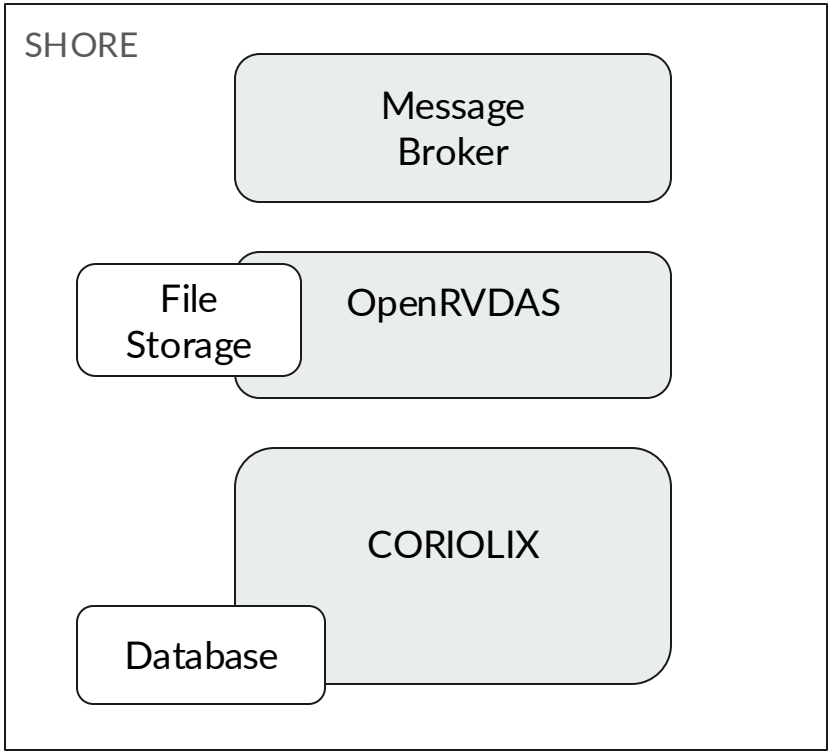
- Site Count = 2 (1 ship, 1 shore)
- Servers/Site = 2 (1 OpenRVDAS, 1 CORIOLIX)
- Total Servers = 4 (2x2)

- OpenRVDAS Server Spec.
  - OS: Linux (many, most?)
  - CPU: 6 (range = 1-24)
  - RAM: 8 (range =8-12GB)
  - DISK: 200GB (range=100 - 500GB)

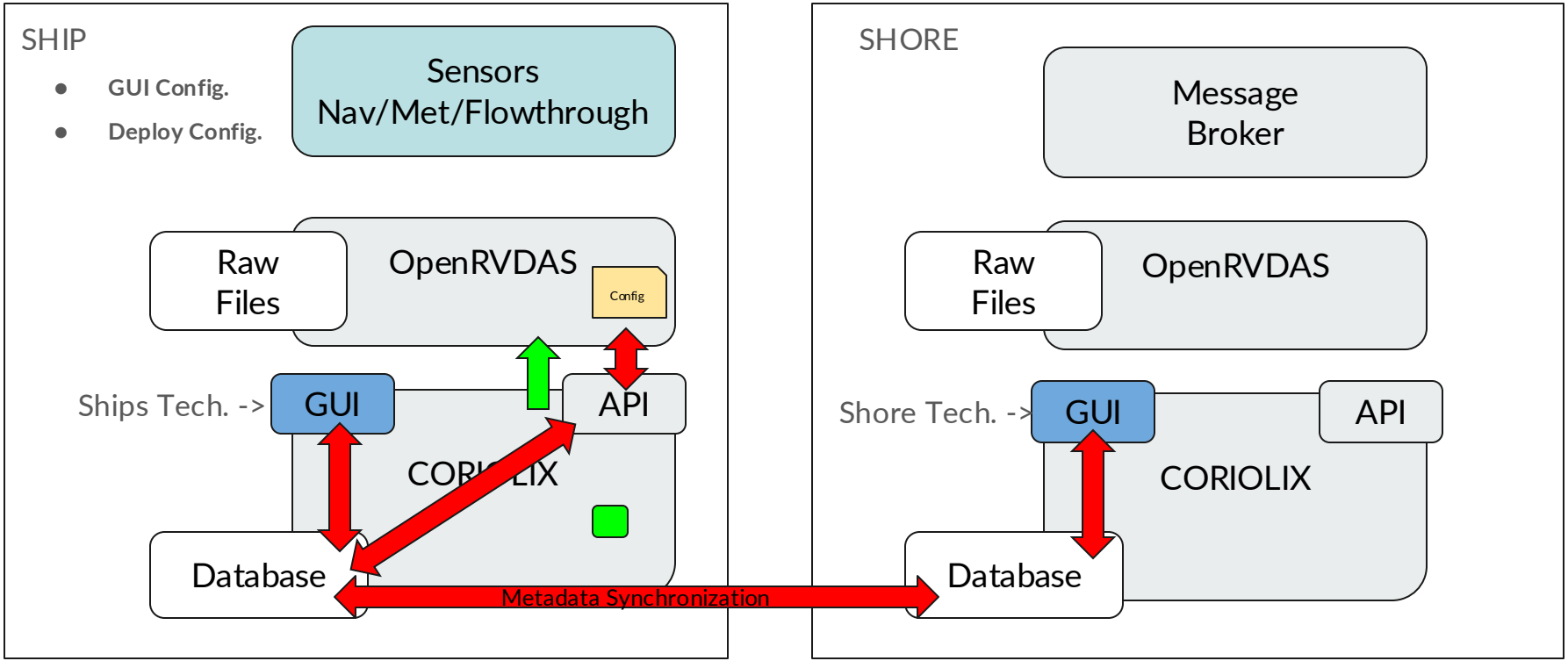
- CORIOLIX Server Spec.
  - OS: Linux (Debian, Ubuntu, Alma)
  - CPU: 12 (range = 1-24)
  - RAM: 24 (range = 12-60GB)
  - DISK: 1TB (range = 250-2TB)

How does it work?
System Components

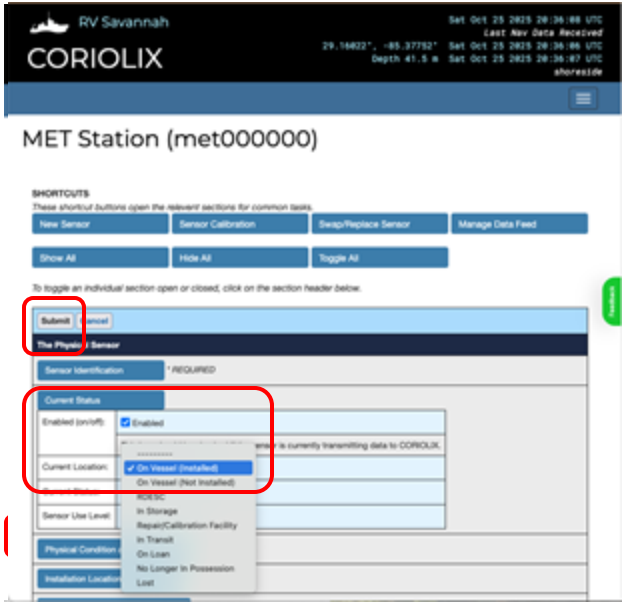




How does it work?
Logging Configuration Management

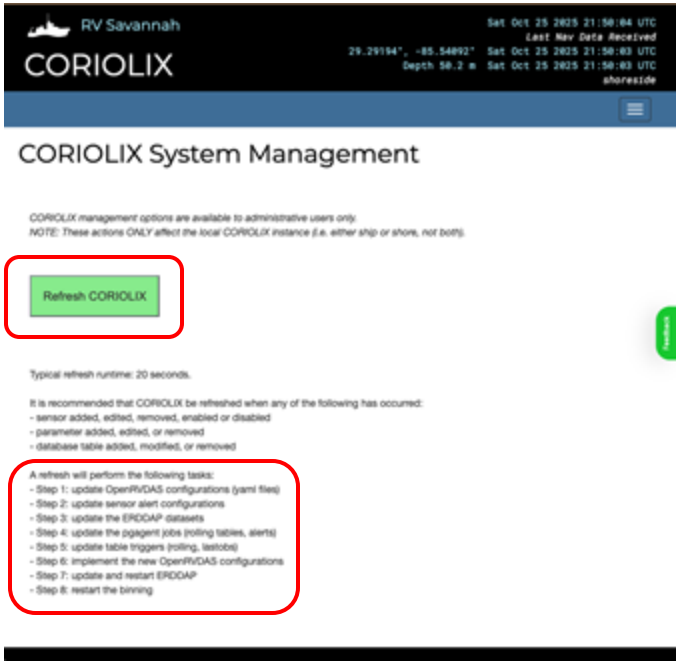


Logging Configuration Management: Part 1



- Edit from ship or shore through the web user interface.
- Login, only technicians can edit & all edits logged.
- Select sensor to edit from your inventory.
- Update the configuration setting.
- Submit to save configuration change locally and update remote.

Logging Configuration Management: Part 2



Deployment Demonstration

- Technician (priv. user) uses system configuration page.
- Depress the BGB
- This kicks off a series of processes:
  - Create OpenRVDAS config.
  - Update alerts
  - Update ERDDAP
  - Update jobs
  - Update triggers
  - Implement OpenRVDAS config.
  - Restarts

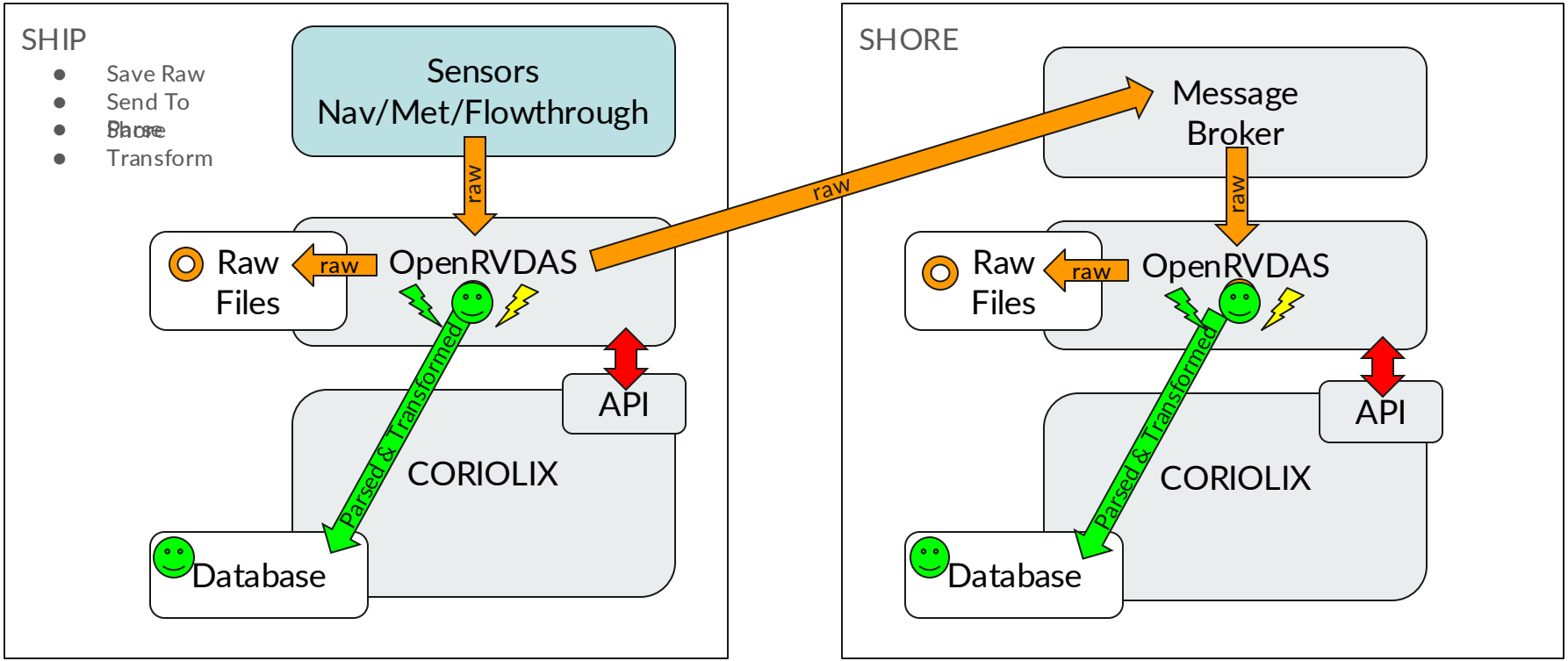
Logging Configuration Management: Part 3

API GET REQUEST: https://coriolix.savannah.skio.uga.edu/api/sensor/?enabled=true&format=json
API GET REQUEST: https://coriolix.savannah.skio.uga.edu/api/parameter/?sensor_id=met000000&format=json

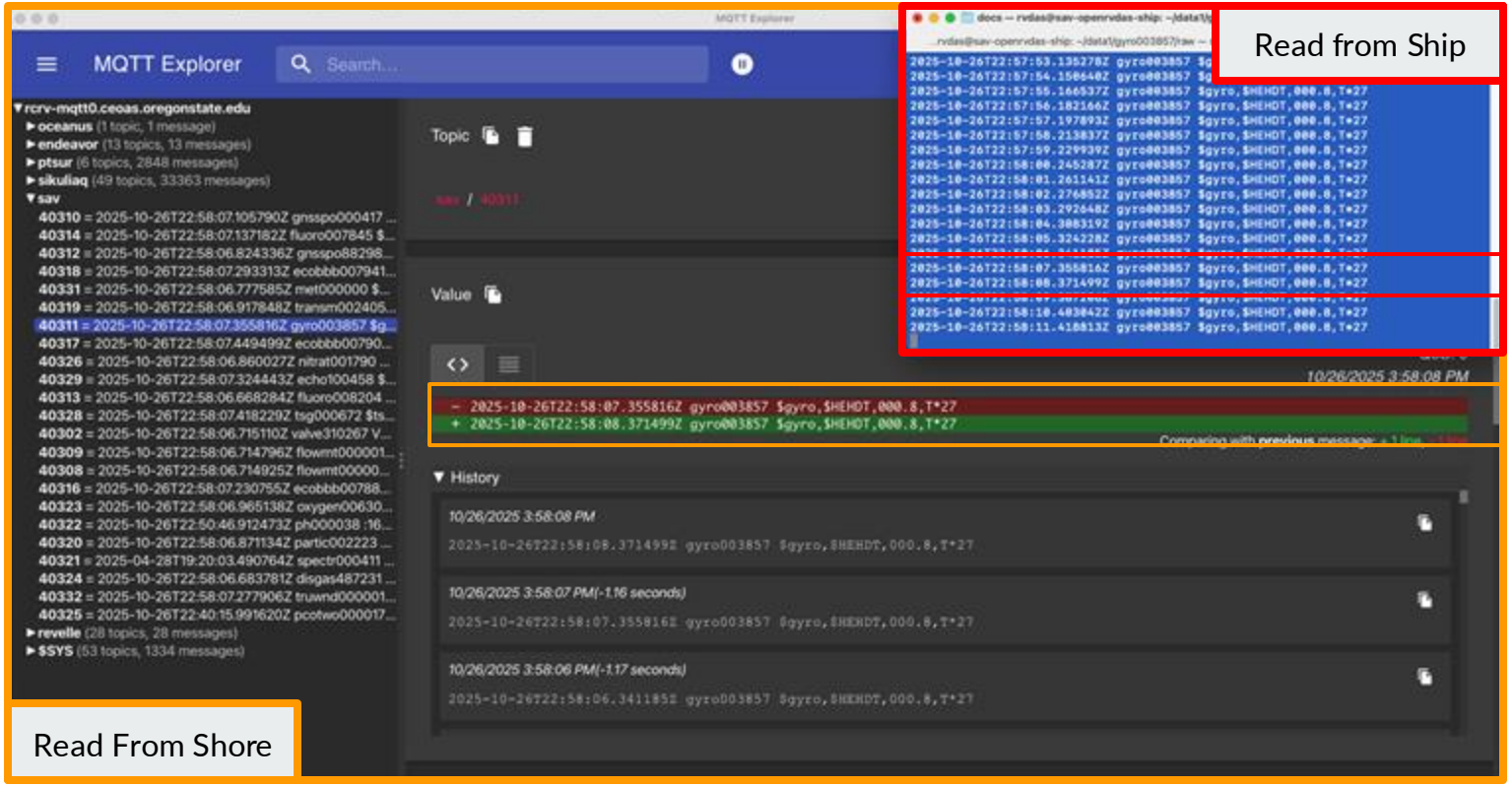
Create OpenRVDAS Config

- BGB Signals OpenRVDAS
- OpenRVDAS queries CORIOLIX API for:
  - Complete set of “enabled” sensors
  - And for each sensor, metadata to:
    - Configure a reader (com. settings)
    - Configure desired transforms
      - Parse sensor messages
      - Ancillary sources
      - Calibration coefficients & transform code
    - Configure desired writers
- OpenRVDAS processes query response using template to produce an OpenRVDAS configuration.yaml file.

How does it work?
Data Flow



DataFlow Part 1: Save & Send



Data Flow Part 2: Parsing

Sensor message parsing is a key step in the data flow, implemented with regular expressions.

What is RegEx? *A regular expression, or regex, is a special text string used to describe a search pattern for matching and manipulating text. Using metacharacters and operators, it can match patterns instead of just literal text, allowing for powerful and flexible text processing.*

Message: $furuno-gga,$GPGGA,002429.00,3159.4030,N,08101.3730,W,1,11,1.4,21,M,,M,,*43

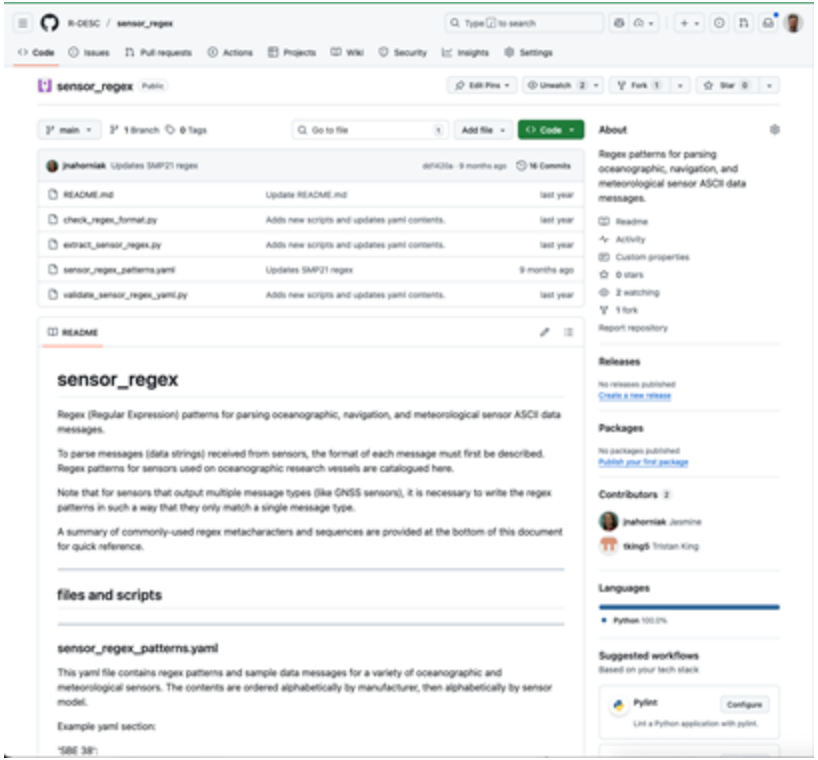
- Utc_position_fix: '002429.00'
- Latitude: '3159.4030'
- Latitude_dir: 'N'
- Longitude: '08101.3730'
- Longitude_dir: 'W'
- Gps_quality_indicator: '1'
- Num_sat_vis: '11'
- Hdop: '1.4'
- Ortho_height: '21'
- Geoid_separation: None
- age: ''
- checksum: '43'

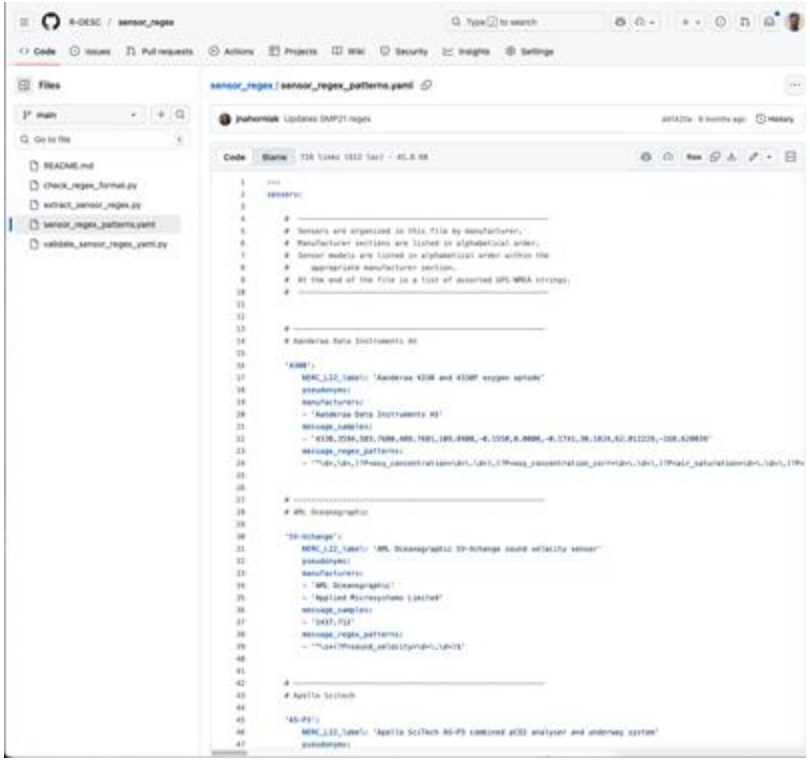
```
^\Wfuruno-gga,
\WGPGGA,
(?P<utc_position_fix>\d+\.\d+),
(?P<latitude>\d+\.\d+),
(?P<latitude_dir>[NS]),
(?P<longitude>\d+\.\d+),
(?P<longitude_dir>[EW]),
(?P<gps_quality_indicator>\d+),
(?P<num_sat_vis>\d+),
(?P<hdop>\d+\.\d+),
(?P<ortho_height>\-?\d+\.?\d*),
M,
(?P<geoid_separation>\-?\d+\.\d+)?,
M,
(?P<age>[\d\.]*)?,
\*(?P<checksum>[0-9A-F]{2})$'
```

Data Flow Part 2: Parsing

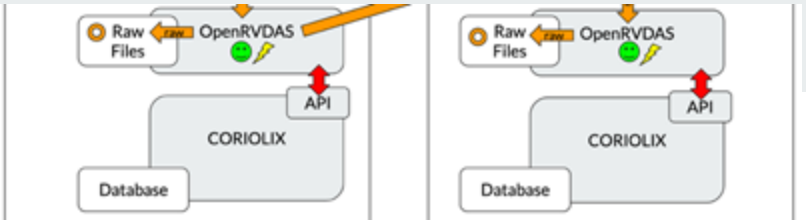
Show CRUISE.yaml here

Data Flow Part 2: CORIOLIX RegEx Repo!





Data Flow Part 3: Transforms




```
tsg000672->prefix/file/db:
  name: tsg000672->prefix/file/db
  interval: 0
  readers:
  - class: UDPReader-
  transforms:
  - class: PrefixTransform-
  - class: TimestampTransform
  writers:
  - class: LogfileWriter-
  - class: ComposedWriter
    kwargs:
      transforms:
      - class: RegexTransform
        module: local.coriolix.logger.transforms.regex_transform
        kwargs:
          field_patterns:
          - ^\Wtsg000672,t1=\s*(?P<temperature>\-?\d+\.\d+),\s*c1=\s*(?P<conductivity>\d+\.\d+),\s*s=\s*(?P<salinity>\d+\.\d+)$
      - class: CalibrationTransform
        module: local.coriolix.modules.calibration_transform
        kwargs:
          sensor_id: tsg000672
      writers:
      - class: CORIOLIXWriter
        module: local.coriolix.logger.writers.coriolix_writer
        kwargs:
          data_table:
            sensor_float_8_archive:
            - p2
            - f2
            - p1
            - f1
            - p3
            - f3
```

calibration_transform.py

- metadata_utilities.py
  - Provides the logic to lookup calibration coefficients & other required metadata
- process.py
  - Imports sensor specific transformation routines

2025 Major Update

- What?
  - **API and Backend Overhaul**
  - Plotting Updates
  - Metadata Repository
  - Shipboard Message Broker & Websockets

Data Model Change

- Parameter Based Storage
  - **OLD - Sensor Tables**
    - sensor_mixlg_11

id	time	time_raw	sensor_id	p1	f1	p2	f2…
42	2025-11-4T04:01:34	2025-11-4T04:01:34	seapth000000	1.123	2212122	5.567	2222122

  - **NEW- Parameter Tables:**
    - latitude_degrees_seapth000000_fullres

id	time	time_raw	flag_summary	flag_string	value
42	2025-11-4T04:01:34	2025-11-4T04:01:34	1	2222122	5.567

Data Model Change

- Simplified Binned Data Structure
  - { "a": [359.017, 5.259, 0.1, 359.9, 120, 359.5, -999], "b": [358.982, 5.321, 1, 359.9, 117, 359.5, -999], "c": [358.982, 5.321, 1, 359.9, 117, 359.5, -999], "fa": "22412222222222222222222222222", "fb": "22112222222222222222222222222", "fc": "22112222222222222222222222222", "sa": "2021-05-06 22:12:59Z", "sb": "2021-05-06 22:12:59Z", "sc": "2021-05-06 22:12:59Z" }
  - { "id": 116331, "time": "2025-10-24T04:00:00Z", "latitude": 57.19606565, "longitude": -146.963442733333, "cog_degreestrue_seapth000000": 168.5911, "sog_kmperhr_seapth000000": 0.305, ...}
- Dedicated Bin Statistics Tables
  - {"id": 8800, "time": "2025-08-04T04:00:00Z", "flag_summary": 1, "flag_string": "22212222222222222222222222222", "spot_time": "2025-08-04T04:00:00Z", "spot_value": 58.0196863666667, "mean": 58.0196873725, "stddev": 8.3127426164e-7, "minimum": 58.0196859666667, "maximum": 58.0196893666667, "median": 58.0196872916667, "num_values": 60}

Data Model Change

- Descriptive Data Naming
  - sensor_mixlg_11
  - latitude_degrees
- Descriptive API Endpoints
  - https://coriolix.example.edu/api/**sensor_mixlg_11**/
  - https://coriolix.example.edu/api/data/**fullres/seapth000000/latitude_degrees**/
- CF Standard Naming
  - ALATSP01
  - Latitude_degrees
- Primary Parameter Designations
  - Adds “is_primary” field to flag parameter as primary data source
  - GUI Page for making and updating primary assignments

Resources:

- https://github.com/R-DESC/CORIOLIX_notebooks
- https://github.com/R-DESC/sensor_regex

CORIOLIX Acknowledgements

Regional Class Research Vessel Program- Oregon State University

Datapresence Student Employees
Shivani Wanjara
Sean Marty
Matthew Zakrevsky
Jack Stevenson
Ian Black
Anna Hughes
Shobana Chandrasekaran
Kyle Buffington
Mina M. Ho

Datapresence Student partners
Hannah Hadi

Datapresence Contractors
Steve Foley
Webb Pinner
David Pablo Cohn

Vessels & Crew
Oceanus
Endeavor
Point Sur
Sikuliaq
Savannah

OSU
Kyle Cole (SMILE)
Renee O'Neill (SMILE)
Adam Talamantes (SMILE)
Amy Malozzi (SMILE)
Tracy Crews (Oregon Sea Grant)

OSU CEOAS
OSU Ship Operations
Research Computing Support
Reimers Lab
Angel White
Maria Kavanaugh
Martechs:
Andrew Woogen
Jonah Winters
Kristin Beem
Brandon D'Andrea
Alex Wick
Emily Shimada
Kate Kouba
Croy Carlin
Michael Tepper-Rasmussen
Sam Richards
Tyler Peterson
Joseph Soltis

NSF Major Facilities
OOI (OSU)
TrustedCI
CI Compass

R2R/SAMOS
Suzanne Carbotte (R2R)
Dru Clark (R2R)
Rebecca (Becca) Hudak (R2R)
Suzanne O'Hara (R2R)
Chris Olson (Scripps)
Karen Stocks (R2R)
Laura Stolp (R2R)
Shawn Smith (SAMOS)
Kristen Briggs (SAMOS)
Mark Bourassa (SAMOS)
Ethan Wright (SAMOS - student)

RCRV OIs
Lynn Butler (ECOC, URI)
William Fanning (ECOC URI, RCRV SOC)
Erich Gruebel (ECOC URI)
Gabe Matthias (ECOC URI)
Kylie Pasternak (ECOC URI)
Joshua Bierbaum (GCOC USM)
Alex Ren (GCOC LUMCON)

ARF & Other External Collaborators
Julia Hummon (UHDAS)
Toby Martin (UHDAS)
Steven Hartz (UAF)
John Haverlack (UAF)
Steve Roberts (UAF)
Ethan Roth (UAF)
Lee Ellet (SIO)
Vicki Ferrini (LDEO)
Dwight Coleman (URI ISC)
Derek Sutcliffe (URI ISC)
John Meyer (HSN)
Kevin Walsh (HSN)
Gwen Moncoiffe (NERC)
Beiko Maas (MFP)
Alice Doyle (UNOLS)

R-DESC alumni
Katie Watkins-Brandt