

NDSF Data & Software Initiatives Update

Tina Haskins - NDSF Associate Director for Data and Science Ops
Vicki Ferrini - NDSF Associate Director for Data System Strategy

NDSF NATIONAL
DEEP SUBMERGENCE
FACILITY



General Data Updates



- SubSea Data Collective

- Oct 2025 SOI-NDSF-MGDS “Data at Sea” workshop
- AGU 2025 session + poster
- Paper (ready for submission)
- New Wiki: <https://github.com/subseadata>



- Adding verbiage to NDSF Data Policy to address Commercial Users

- Worked with WHOI legal on clear terms for current use cases
- Will ask DeSSC to review

- Archive Digitization Project

Friendly Reminders

Cumulative Dive Stats

Alvin Dive Stats	
Vehicle Type	HOV
Number of Expeditions	703
Number of Dives	5,313
Avg. Max Depth (meters)	-2,075
Total Dive Time (hours)	38,288
Total Time on Bottom (hours)	725
Avg. Dive Time (hours)	7
Avg. Time on Bottom (hours)	5

Jason Dive Stats	
Vehicle Type	ROV
Number of Expeditions	155
Number of Dives	1,773
Avg. Max Depth (meters)	-1,751
Total Dive Time (hours)	25,338
Total Time on Bottom (hours)	20,514
Avg. Dive Time (hours)	14
Avg. Time on Bottom (hours)	12

Sentry Dive Stats	
Vehicle Type	AUV
Number of Expeditions	84
Number of Dives	757
Avg. Max Depth (meters)	-2,225
Total Dive Time (hours)	11,089
Total Time on Bottom (hours)	9,499
Avg. Dive Time (hours)	15
Avg. Time on Bottom (hours)	13

Vehicle Dive Stats

Select Time Range: (Jan 1, 2026 → May 19, 2026)

Jason	
Vehicle Type	ROV
Number of Expeditions	2
Number of Dives	21
Avg. Max Depth (meters)	-2,298
Total Dive Time (hours)	365
Total Time on Bottom (hours)	298
Avg. Dive Time (hours)	17
Avg. Time on Bottom (hours)	14

Sentry	
Vehicle Type	AUV
Number of Expeditions	1
Number of Dives	6
Avg. Max Depth (meters)	-5,556
Total Dive Time (hours)	187
Total Time on Bottom (hours)	160
Avg. Dive Time (hours)	31
Avg. Time on Bottom (hours)	27

Sealog

Welcome to Seaplay

Please select a cruise from the list below. Once a cruise is selected please select a lowering from the list of lowerings associated with that cruise that appear at the bottom of the cruise information panel. Selecting a lowering will open the lowering information panel. At the bottom of the cruise information panel there will be links for proceeding to the lowering replay section, the lowering review section and the lowering gallery section of Sealog. If at any time you wish to return to this page please click the "Review Cruises/Lowerings" text in upper navigation bar.

Year: 2025	Cruise: AT50-42	Dive: Alvin-D5340
AT50-42 AT50-38 AT50-36 AT50-34 AT50-33	Cruise Name: Endeavor N-cycling Under Pressure Chief Scientist: Jill McDermott Vessel: R/V Atlantis Location: NE Pacific Dates: 2025/09/05 → 2025/09/22 Ports: Newport, OR → Astoria, OR Duration: 17 days 12 hours 0 minutes Dives: Alvin-D5340 Alvin-D5339 Alvin-D5338 Alvin-D5337	The data you are trying to view is currently under embargo. For access, please contact: ndsf_info@whoi.edu
Year: 2024		
Year: 2023		
Year: 2022		

- Metadata accessible for embargoed Cruise/Dives
- Sealog database migration planning: MongoDB to PostgreSQL
- Successful deployment of Sealog Sandbox

Request a Sealog Sandbox

Get your own temporary Sealog instance for testing and exploration. We'll email you the access details within moments.

Email Address

Event Templates

JASON Templates

Choose which event templates to pre-load in your sandbox

✓ Success!



Request Sandbox

ALVIN

JASON

SENTRY

WAVEGLIDER

DATA

PLAN AN EXPEDITION

NEWS & MEDIA

List Map

Vehicle

Current NDSF Vehicle

Has Data at MGDS

Has Framegrab Link

Cruise ID

Feature (search string)

Scientist

Pilot

West

East

South

North

Start Date

End Date

Search

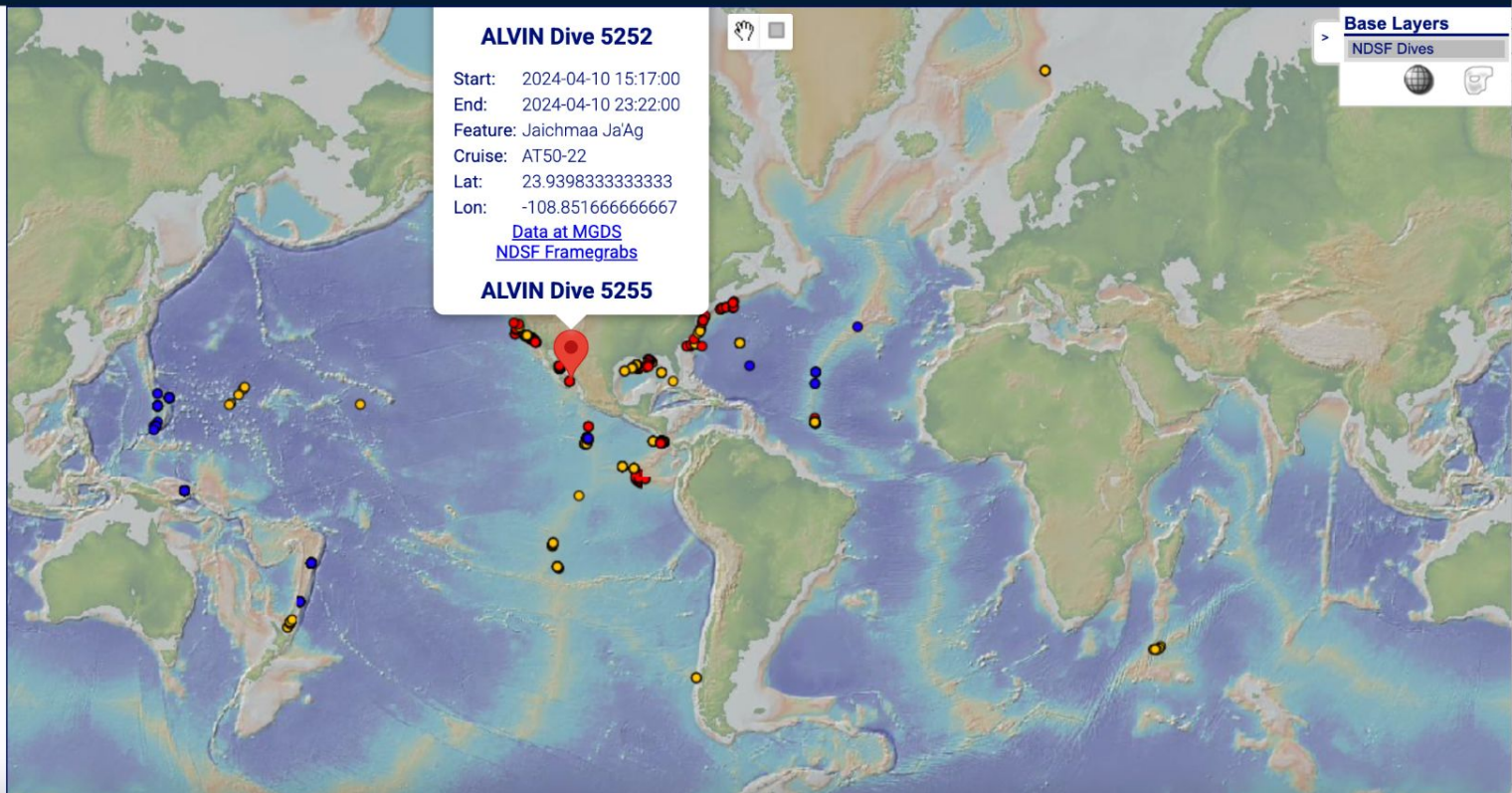
Download Results as:

ALVIN Dive 5252

Start: 2024-04-10 15:17:00
 End: 2024-04-10 23:22:00
 Feature: Jaichmaa Ja'Ag
 Cruise: AT50-22
 Lat: 23.9398333333333
 Lon: -108.851666666667

[Data at MGDS](#)
[NDSF Framegrabs](#)

ALVIN Dive 5255



Accelerating NDSF Data Ingest for FAIR Access at MGDS

Historical Challenges:

- Fragmented metadata
- Heterogeneous data
- Legacy media formats
- Manual MGDS file set curation requiring deep understanding of NDSF operations and practices

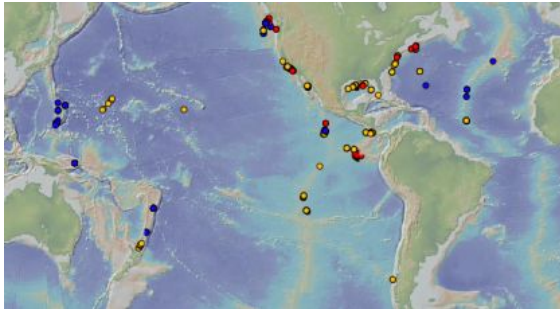
The Solution:

Semi-automated Python pipeline to generate metadata files that feed directly into MGDS ingestion code:

- API & lookup queries harvest metadata from R2R, WHOI DLA, MGDS; operator verifies
- Pattern matching detects data file presence
- Cruise addenda record file modifications

What's Next:

- Systematic processing of historical backlog
- Additional cross-vehicle standardization
- New vehicle support
- Addendum annotations
- Enhanced download statistics



```
SAMPLE_cruise_config.yaml
entry_id: AT99-99
dms_release_date: 2025-12-14
investigators:
  person_id: Livingston_Kingston
  institution: edu.sesame
Platform_folder: Atlantis
data_sets:
  suffix: scc
```

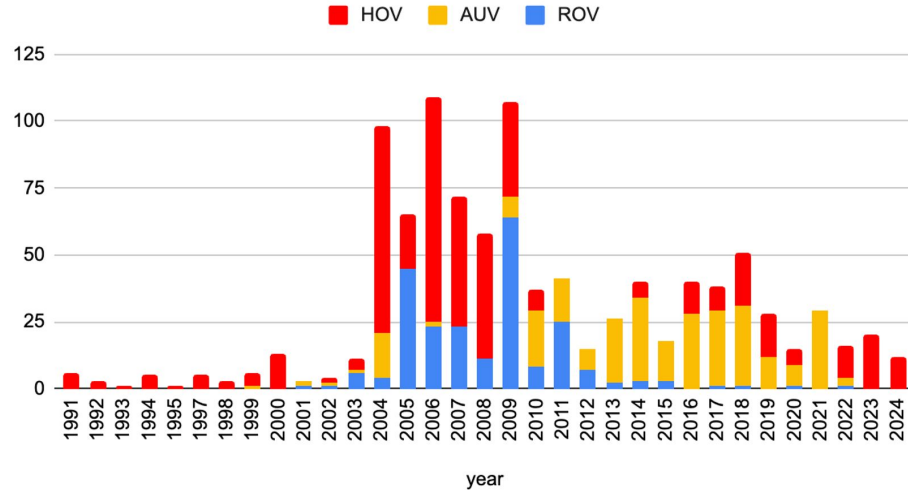
Special Thanks to:

Catie Graver WHOI/NDSF
John Morton LDEO/MGDS

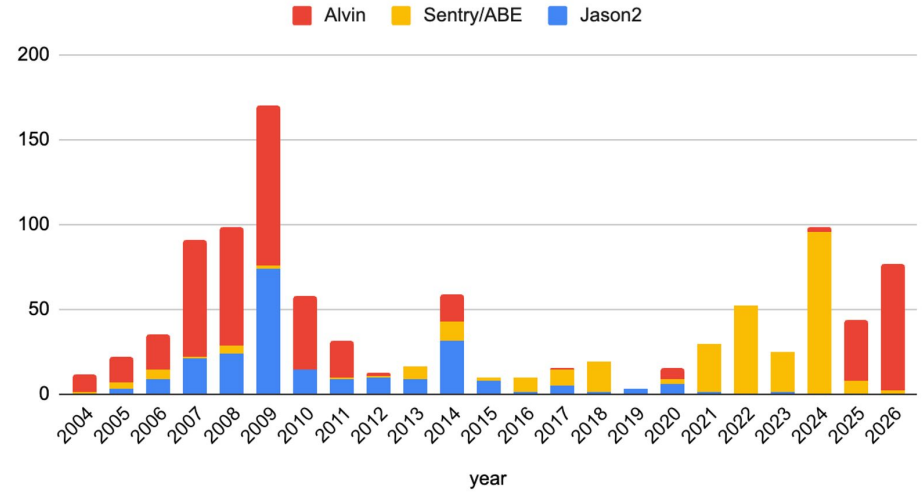


Accelerating NDSF Data Ingest for FAIR Access at MGDS

NDSF Data sets in MGDS by Year of Acquisition

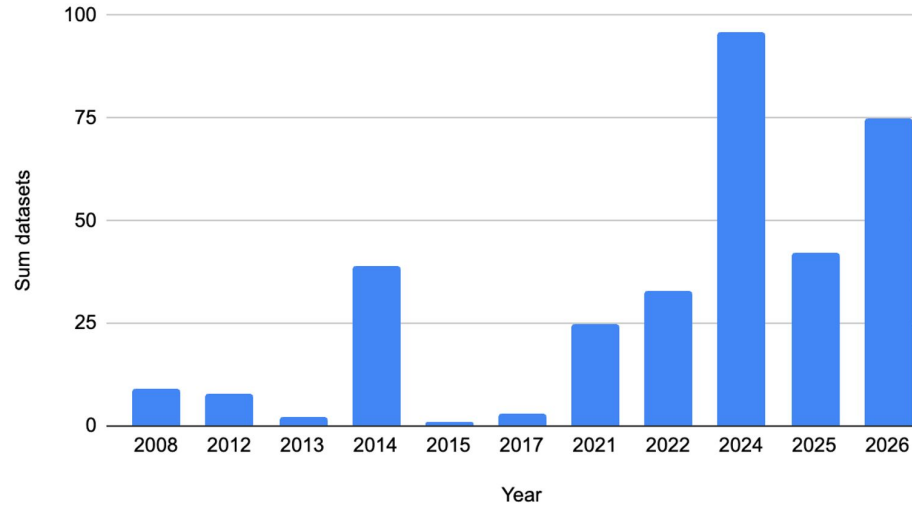


NDSF Data Sets at MGDS by Year of Ingestion and Vehicle

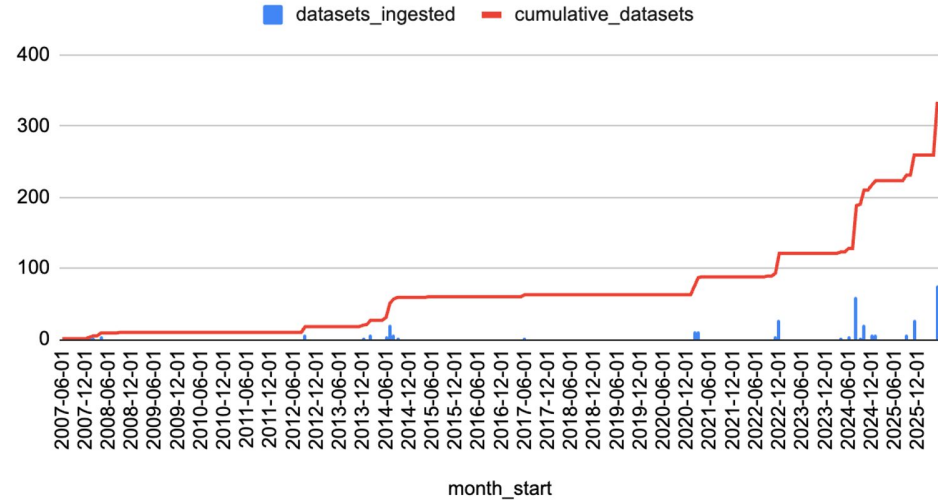


Accelerating NDSF Data Ingest for FAIR Access at MGDS

Total NDSF datasets in MGDS by year of ingestion

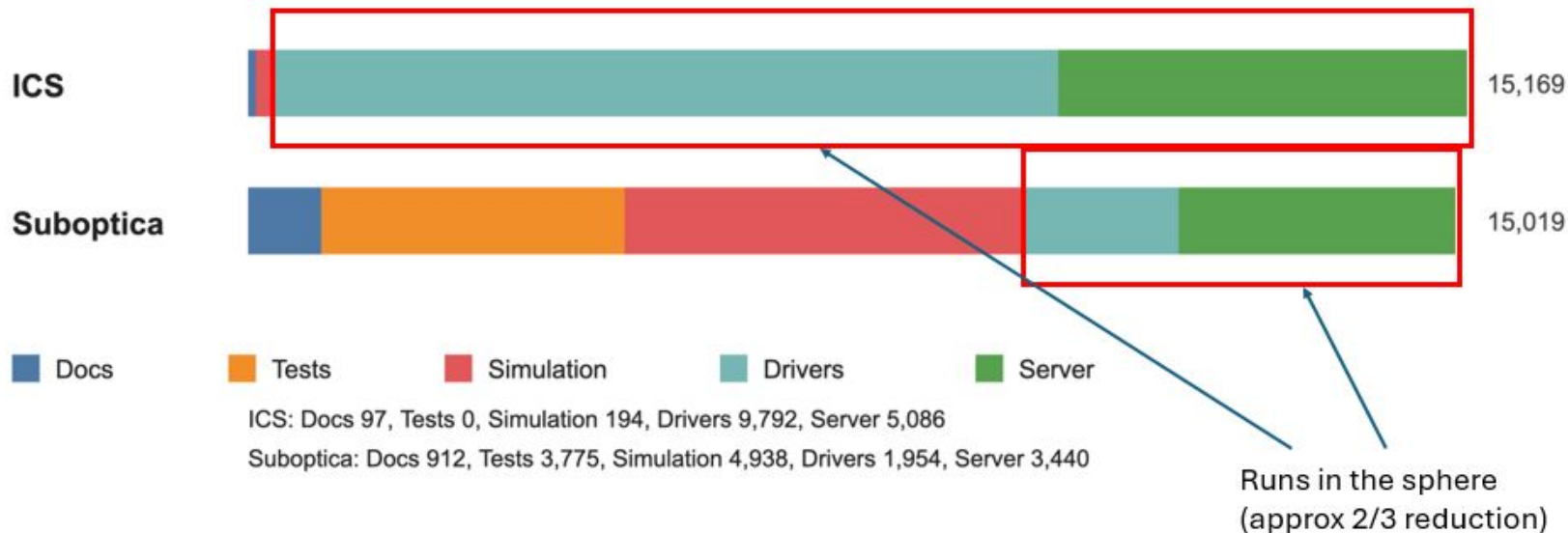


Monthly Ingestion of Data Sets contributed by NDSF



Alvin AIS Update

Lines of Code Change

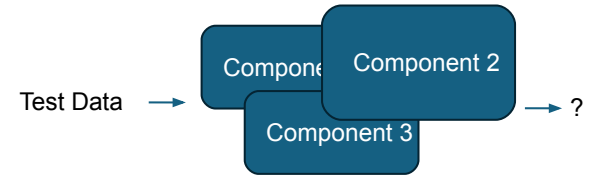


AIS Update Continued

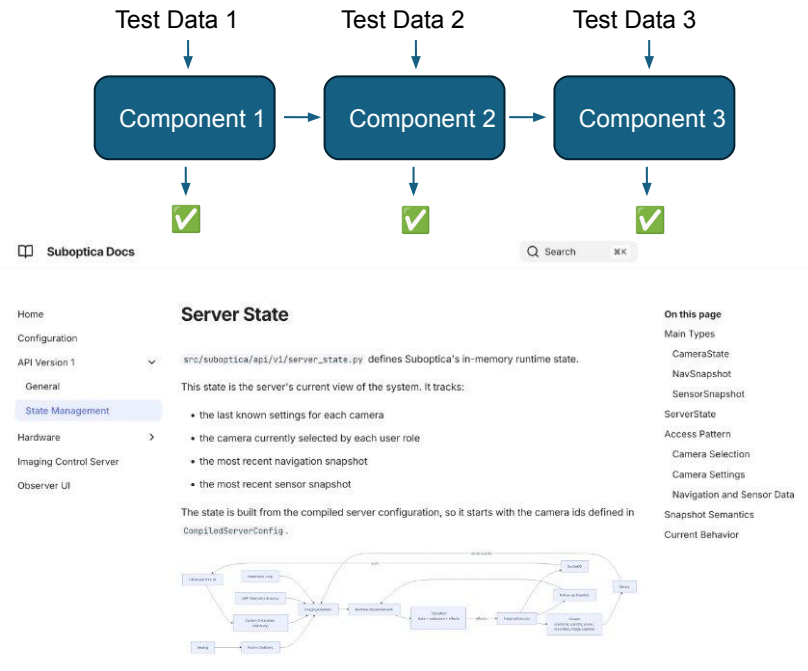
Anticipated Benefits

- Error handling + Observability
 - Capturing what happens in the sphere takes the guesswork out.
 - Identified and resolved single drive failure modes
- Tests
 - Unit-level testing enforces clean interfaces
 - 96% test coverage so far
 - Makes code changes and bug fixes safer to deploy
- Continuous Deployment
 - Regular practice, not an arcane ritual
- Documentation
 - In process of making this easier for the new devs to understand

Without clean interfaces – unknown internal state



With clean interfaces – can be tested in isolation



Codebase Evolution - *Science Impact at Each Step*

2025

Platform reliability

Alvin C&C stability
Jason navigation fixes
Sentry ROS node improvements
DSLPP pipeline rollout

Reclaiming dive time with cleaner data

2026

Cross-vehicle ROS 2 codebase

Alvin, Jason, Sentry, REMUS unified
Plug and play integration for new science sensors
Self-describing data formats
Remote debugging tools

Cross-platform data registration & structured datasets

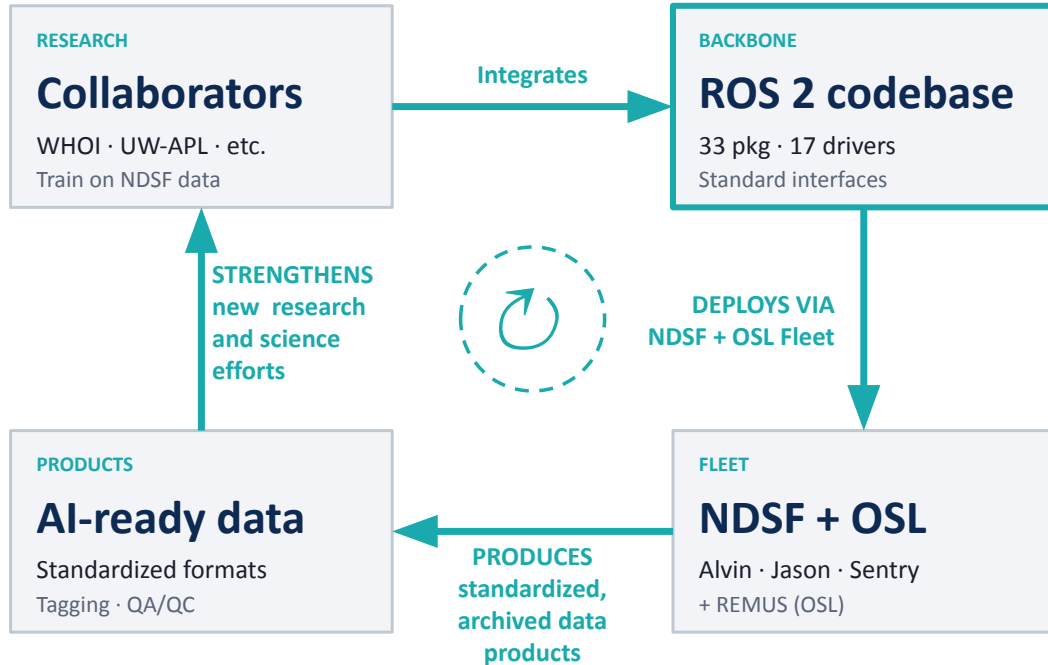
2027+

Building on top

WHOI mission executive
ATLAS archive synthesis
AI/ML data pipelines
Cross-WHOI tooling

An interactive data archive & AI becomes operational

ROS 2 Migration - Progress



The substrate is built.

Core packages: *Ready*

Sensing: *Ready*

Navigation: *In Development*

Control: *In Development*

Per-vehicle progress.

Alvin Testing underway → June sea trials

Jason Development → Eng Dive in Fall

Sentry Development → Maint. Period EOY

For the first time, the same software stack runs across all WHOI vehicles:
making cross-platform comparison at the same site reproducible

ATLAS - AI-Enabled Tool for Longitudinal Archive Synthesis

ATLAS points to existing records. It does not invent new science data.

Expert verification at every step is a design requirement, not an afterthought.

01 · Chatbox Interaction

Ask in natural language

The scientist poses a research question, not a directory search

EXAMPLE QUERY

"Compare salinity from dive 001 and dive 002 on the same plot."

02 · ATLAS RESOLVES

The system shows its work

Pre-defined tools find, analyze, and plot existing data

- 01 **list_sensors(dive_id)**
Sensors, timestamps, vehicle.
- 02 **inspect_data_quality(series[])**
Verify alignment, gaps, and sampling before any comparison is made.
- 03 **plot_multi(series[])**
Compare the selected existing series; it does not generate new science data.

03 · SCIENTIST INSPECTS

Trace back to existing records

Loaded data includes a full provenance chain

- **dive_001 / salinity**
Existing time-series stream loaded from the archive and passed to plot_multi(series[]).
- **dive_002 / salinity**
Existing time-series stream loaded from the archive and compared against dive_001.
- **provenance fields**
expedition · vehicle · dive · sensor · timestamps

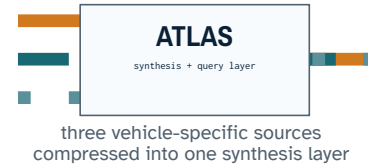
SAME RAW INPUTS

sensor records
nav data
dive logs

TODAY



WITH ATLAS



VALIDATION DATASET

2021–2026 NDSF data

8

EXPEDITIONS

Alvin: 6 packages Sentry: 4 packages Jason: 2 packages

12

VEHICLE PACKAGES

~62 TB

EPR DATA



Scientists burn ship time wrangling data instead of maximizing expedition objectives.

2021–2026 captures the run-up to and aftermath of the 2025 eruption.

Questions/Feedback

- Reach us at ndsf_info@whoi.edu