

Developing a New Vision for NDSF Data Management

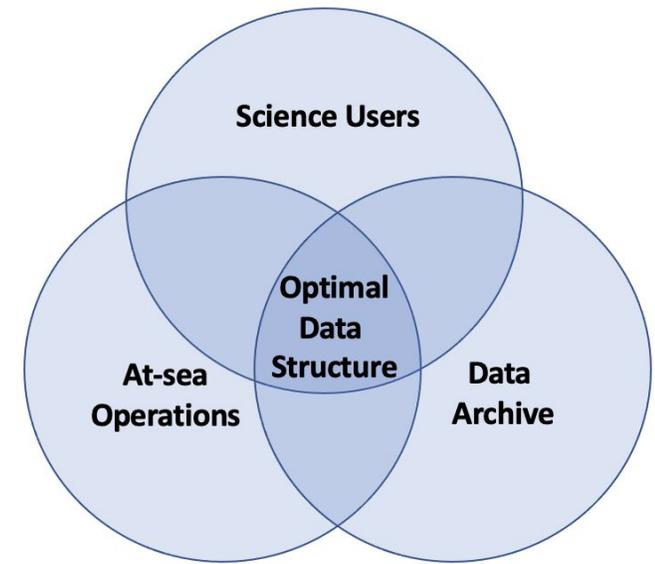
Vicki Ferrini & Tina Haskins

NDSF NATIONAL
DEEP SUBMERGENCE
FACILITY



Vision

- Develop a new strategy for the management and delivery of NDSF data that meets the needs of at-sea scientists and operations teams and down-stream data repositories that serve the broader science community and public
- Ensure that data management activities, and the maintenance and development of software that supports data management and access, are coordinated and are moving toward a common vision
- Standardize data structures and at-sea protocols to gain efficiencies and achieve cross-vehicle consistency
- Optimize the management, versioning, and maintenance of software tools



New NDSF Data Team

Software
Team

Data Archiving
and Access Team

Associate
Data Director

Vehicle Data Staff



- Vicki Ferrini: NDSF Data Director
 - Vision & Strategy
 - Expertise: Shoreside Data
- Tina Haskins: NDSF Associate Data Director
 - **NDSF Data POC - ndsf_info@whoi.edu**
 - Expertise: At-Sea Data
- Stefano Suman: Software Lead
 - Expertise: Acquisition and Systems

Year 1 Goals

Document & review current at-sea practices

At-sea SOPs, metadata, data directory structures, naming conventions.

Identify vehicle-specific successes, bottlenecks and challenges

Leverage successes and strive toward facility-wide solutions.

Develop cross-vehicle guidelines

Where possible, develop cross-vehicle standardized directory structure, file naming conventions, formats, and generating machine-readable metadata.

Collaboratively design and implement a new structure

Improve efficiencies for at-sea operators and scientists, shoreside data managers, standardize data distributions and optimize data and metadata packages.

Long-Term Goals

- Improve shore-side data management to broaden access
- Work with WHOI Data Library to seek efficiencies and data redundancy
- Leverage efforts within MGDS for submersible data discovery and access
- Make data available for next-generation processing and integration (e.g. AI/ML)
- Curate NDSF data holistically as a coherent collection of data
- Address existing NDSF data that is not yet directly accessible to the community
- Make NDSF data resources available and interoperable with other similar data acquired by other groups both domestically and internationally
- Leverage evolving community solutions for video archiving and access

Submersible Data at MGDS

- Submersible data from 21 vehicles
 - Total: 6 TB, > 1 million files, 1,400 dives, 137 cruises
 - NDSF: 1.6TB, >160k files, 830 dives, 85 cruises
- Data Types:
 - Dive metadata
 - Dive logs and reports
 - Bathymetry, Magnetics, Sidescan, Navigation, Subbottom, Interpretations, Heatflow, Turbidity, Eh, Photomosaic, Photos, FrameGrabs, KMLs, (including Sealog), and more...
 - Virtual Van and FrameGrabber APIs/UIs
- User Interfaces for discovery and self-service
- Web services for interoperability

MGDS
MARINE GEOSCIENCE DATA SYSTEM

About Tools Data Collections Resources Search Data

Search Data Contribute Data Web Services

Submersible Data
Submersibles provide important high-resolution geophysical data sets as well as bottom samples, photos and video that are important for seafloor characterization and ground-truth. Data have been acquired by the National Deep Submergence Facility (NDSF), the Schmidt Ocean Institute (SOI), and the Monterey Bay Aquarium Research Society (MBARI).

- Search for ROV data (e.g. Jason and SuBastian)
- Search for AUV Data (e.g. Sentry and MBARI Mapping AUV)
- Search for HOV Data (Avin)

Data Download Report for Sentry
January 01, 2020 to December 31, 2020

Data Download Reports are prepared bi-annually and are sent to all Scientists associated with downloaded data including Chief/Co-Chief Scientists of field programs, project PIs and Co-PIs, and data set Investigators. Please [contact us](#) with any questions or concerns.

MGDS Data Downloads

Collection ID	Chief Scientist	Data Set Investigator(s)	Device Info	Total Size	Total File Downloads	Data Set	Download Purpose(s)
AT15-03	Sinton	Sinton, White	AUV: Sentry Sonar:Multibeam	586 MB	18	Bathymetry (Grid) doi: 10.1594/EDA/318210	
AT18-03	Fisher	Fisher	AUV: Sentry Sonar:Multibeam	301 MB	2	Backscatter Acoustic Bathymetry Swath (Swath) doi: 10.1594/EDA/320788	
AT18-03	Fisher	Fisher	AUV: Sentry Sonar:Multibeam	734 MB	3	Backscatter Acoustic Bathymetry Swath (raw, Swath) doi: 10.1594/EDA/320804	
AT15-53	Valentine	Valentine	AUV: Sentry Sonar:Multibeam	3 GB	15	Backscatter Acoustic Bathymetry Swath (raw, Swath)	
AT15-53	Valentine	Valentine	AUV: Sentry Sonar:Multibeam	10 kB	1	Backscatter Acoustic Bathymetry Swath (Swath) doi: 10.1594/EDA/321436	

Sealog Updates

- **Code consolidation** - normalized versions used across vehicles
- **Containerization** to facilitate and simplify deployment
- **Continuous integration** - more efficient code validation and testing
- COVID-response: modifications for shoreside watchstander logging
 - Optimize bandwidth use via shore-side caching
- **Back-end modifications** position code base for more rapid development, diagnostics and testing
- New ideas for improving software architecture to gain further efficiencies
- Work is on-going and will be overseen by NDSF Data Leadership

The image shows two overlapping screenshots. The top one is a dark-themed 'System Management' interface for 'Sealog for JASON v1.0.12-rc'. It features a 'Welcome to Sealog' message and instructions for selecting a cruise and lowering. Below this is a vertical sidebar with years from 2011 to 2021. The bottom screenshot is a GitHub repository page for 'WHOIGit / ndsf-sealog-server'. It shows the repository's file structure, including folders like .github/workflows, config, docs, lib, misc, plugins, routes, and test, and files like .eslintrc, .gitignore, and CHANGELOG.md. The repository has 232 commits and 14 tags.

NDSF Website



- Currently under renovation
- Feature updates
 - Training videos
 - Documentation
 - Newsletter Archive
- NDSF Newsletter Sign Up

<https://ndsf.whoi.edu/sign-up-for-the-ndsf-newsletter/>

- Questions/Suggestions? Please contact us via ndsf_info@whoi.edu

Data acquired with NDSF vehicles are valuable digital resources that enable scientific exploration by not only cruise participants but to a diverse community of scientists and educators. Recognizing the importance of these resources, NDSF is re-envisioning its data management practices to establish a modern, scalable, and sustainable data infrastructure that meets the needs of at-sea vehicle operations teams and sea-going scientists as well as shore-side data managers and the broader community.

The initial goals of this work are to review current data management practices to identify challenges and opportunities and to contribute to the development of an NDSF-wide vision and technical solution that will yield efficiencies while meeting the needs of stakeholders. Once optimizations related to new data acquisition, attention will be turned toward improving access to historic data holdings.

To formulate this new vision and make tangible steps toward a new framework for NDSF Data Management, a Data Group has been established within the facility. Led by NDSF Data Director Vicki Ferrini (LDEO) and NDSF Associate Data Director Tina Haskins, this group includes software engineers and sea-going data personnel as well as shoreside data management and geoinformatics specialists. Community input will be solicited at various stages during this process, and progress will be reported in NDSF newsletters and during DESSC meetings. Questions and comments can be directed to ndsf_info@whoi.edu.

NDSF Vehicle Data

Vehicle program-specific data, initially collated by each vehicle program, can be found by clicking on the below icons.

(Note: This data is undergoing metadata quality assessment.)

