

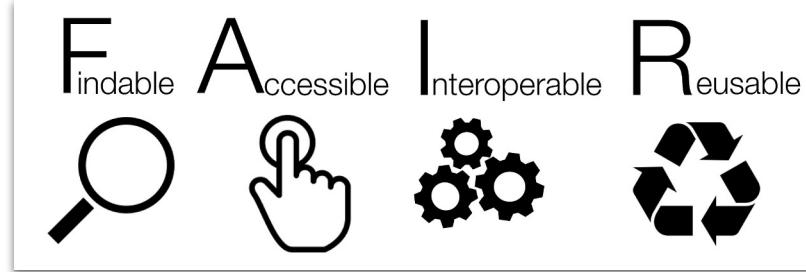
DeSSC New-User Program: Data Management Overview

Vicki Ferrini, LDEO

Open Data Access

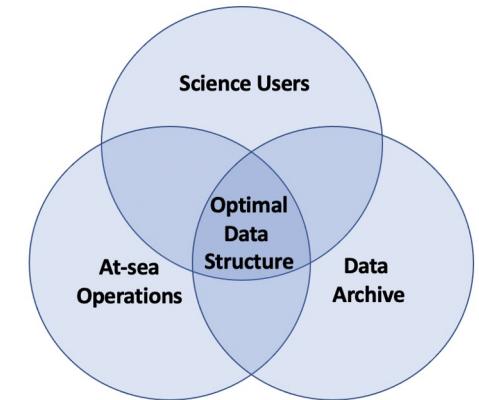
- Driven by:
 - Federal and funding-agency requirements
 - Journal and publisher expectations
 - High acquisition costs that incentivize reuse
- Enables New Opportunities:
 - Tracking spatial & temporal change
 - Scientific transparency and reproducibility
 - Cross-disciplinary data synthesis
 - New analytical pathways, including AI/ML

Open and well-stewarded scientific data continue to be recognized at the national level as essential for advancing research, supporting innovation, and ensuring that publicly funded observations provide the greatest possible benefit to science and society.



Data Accessibility Depends on All of Us

- **Data producers** → share, document, and preserve
- **Data stewards** → curate, maintain, and support discovery
- **Data users** → cite, credit, and contribute improvements
- **Journals & agencies** → reinforce best practices

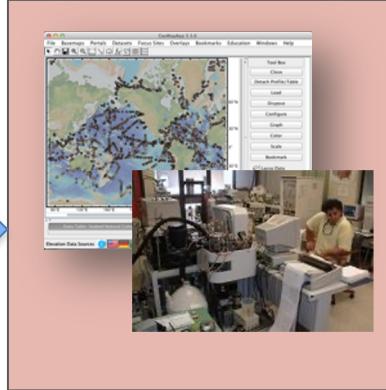
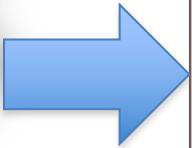


Accessibility is a shared responsibility — and shared benefit.

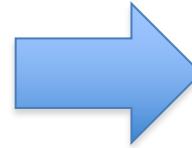
Historic Scientific Workflow



Data Acquisition



Data Processing
& Interpretation



Earth and Planetary Science Letters, 104 (1991) 513-534
Elsevier Science Publishers B.V., Amsterdam

[C1]

Hydrothermal vent distribution along the East Pacific Rise crest (9°09'-54'N) and its relationship to magmatic and tectonic processes on fast-spreading mid-ocean ridges

Rachel M. Haymon ^a, Daniel J. Fornari ^b, Margo H. Edwards ^{b,c}, Suzanne Carbotte ^a, Dawn Wright ^a and Ken C. Macdonald ^a

^a Marine Science Institute and Department of Geological Sciences, University of California at Santa Barbara, Santa Barbara, CA 93106-9500, USA

^b Lamont-Doherty Geological Observatory of Columbia University, Palisades, NY 10564, USA

^c Department of Geological Sciences, Columbia University, Palisades, NY 10564, USA

Received November 27, 1990; revision accepted March 18, 1991

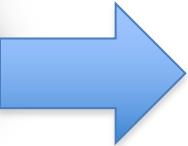
ABSTRACT

Using the near-bottom ARGO imaging system, we visually and acoustically surveyed the narrow (< 200 m wide) axial zone of the fast-spreading East Pacific Rise (EPR) along 83 km of its length (19°09'-54'N), discovered the Venture Hydrothermal Fields, and systematically mapped the distribution of hundreds of hydrothermal features relative to other fast-spreading ridges. The hydrothermal vent distribution is highly variable. The distribution consists of a 2nd order segment and includes at least 4th order (5-15 km) segments defined by bends or small lateral offsets of the ridge crest or axis (Devals). 4th order segmentation of the ridge crest is clearly expressed in the high-resolution ARGO data by the fine-scale behavior of the ridge axis. The distribution of hydrothermal features is controlled by the ridge axis, hydrothermal vent density, and structure, etc. across segment boundaries. The distribution and along-strike variability of hydrothermal features corresponds closely to the morphology and structural segmentation of the ridge. On the 2nd order scale, we find high T hydrothermal activity associated with (1) short segments of axial hydrothermal venting, reflected in the 1.7 km long segments of the ridge axis, and (2) with the presence of a well-developed axial summit caldera (ASC). Previous work refers to this feature as an axial magma chamber (AMC). The presence of an ASC on the 2nd order scale is the primary evidence for the 2nd order scale. There is no evidence that fast-spreading ridges are as elongate volcanic calderas rather than a tecnode graben, and supports the introduction of "axial summit caldera" as a more accurate descriptor. All but 1 of the 45 active high T vent features identified with ARGO are located within the axial zone of the ridge. The distribution of hydrothermal features is controlled by the ASC, no important signs of venting were seen beyond the axial zone. On the 4th order scale, the abundance and distribution of hydrothermal features changes across 4th order segment boundaries. We find that high T vents are most abundant where (1) the AMC is very narrow (40-70 m), (2) the AMC reflector is most shallow (< 155 m) beneath the axial zone, and (3) the axial zone is very narrow (40-70 m), (4) the AMC reflector is most shallow (< 155 m) beneath the axial zone.

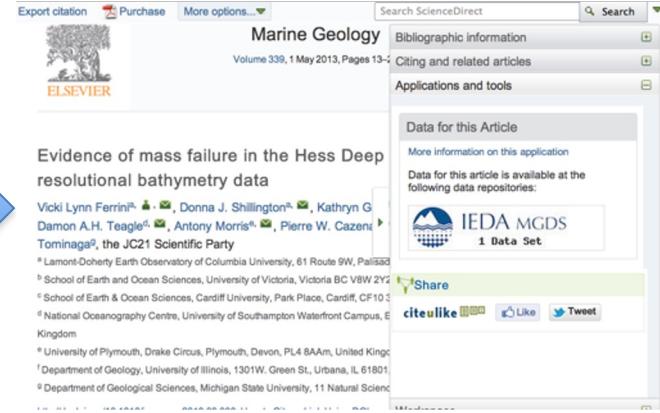
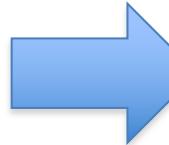
Modern Scientific Workflow



Data Acquisition



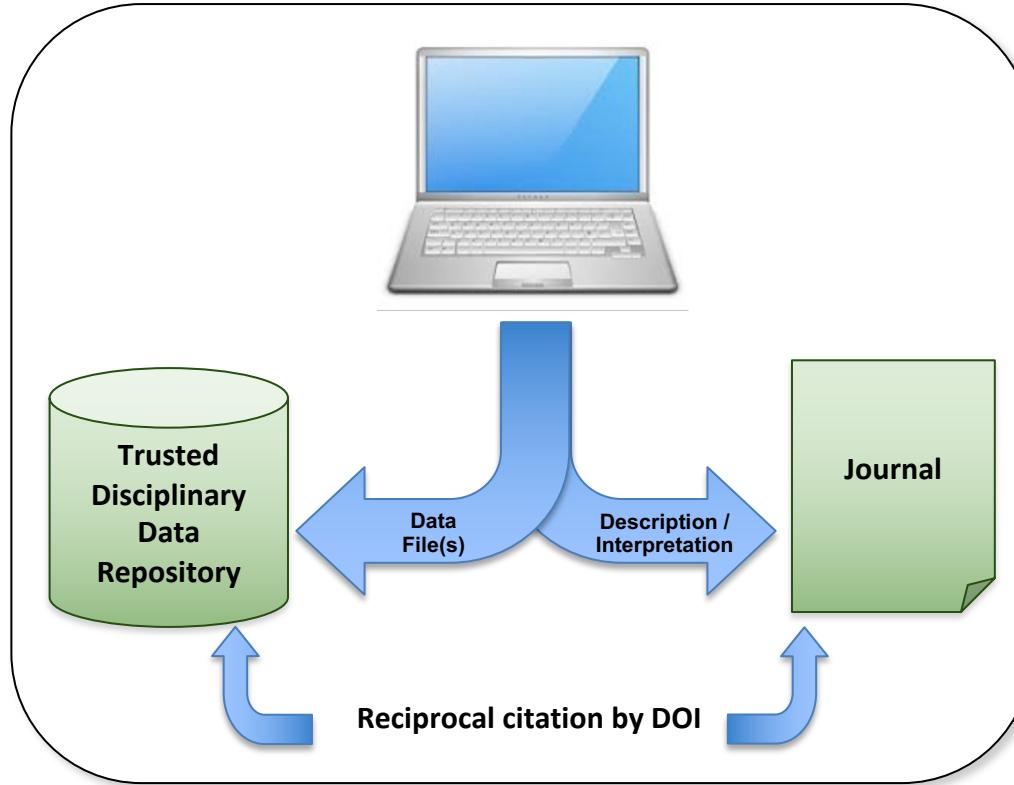
Data Processing
& Interpretation



Publications



“Best Practice”



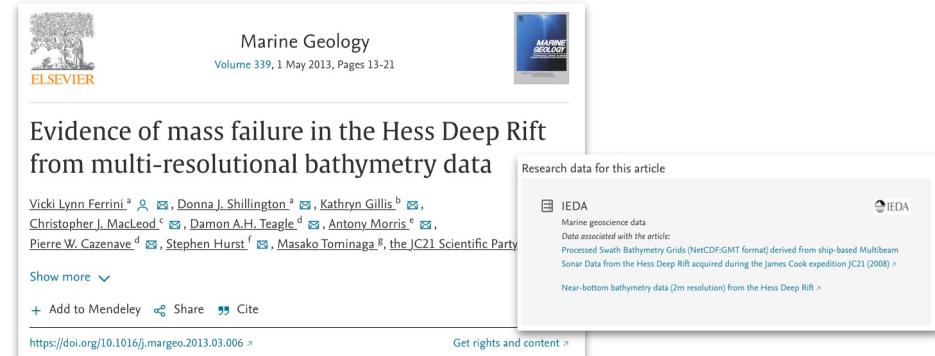
What's in it for you?

- Scientific integrity & reproducibility
- Collaboration
- Research Efficiency
- Increase research impact
- Data citation
- Compliance with data policies
- Outreach & Engagement
- Preserve data for your own future use

“The coolest thing to do with your data will be thought of by someone else.”

Rufus Pollock

Cambridge University and Open Knowledge Foundation



The screenshot shows a journal article from the *Marine Geology* journal, Volume 339, 1 May 2013, Pages 13-21. The article is titled "Evidence of mass failure in the Hess Deep Rift from multi-resolutional bathymetry data". The authors listed are Vicki Lynn Ferrini, Donna J. Shillington, Kathryn Gillis, Christopher J. MacLeod, Damon A.H. Teagle, Antony Morris, Pierre W. Cazenave, Stephen Hurst, and Masako Tominaga, with the JC21 Scientific Party. Below the article title, there is a "Research data for this article" section. It includes a link to the IEDA (Integrated Earth Data Applications) portal, which provides marine geoscience data associated with the article. The data is described as "Processed Swath Bathymetry Grids (NetCDF-GMT format) derived from ship-based Multibeam Sonar Data from the Hess Deep Rift acquired during the James Cook expedition JC21 (2008)." A "Near-bottom bathymetry data (2m resolution) from the Hess Deep Rift" is also mentioned. At the bottom of the article page, there are links to "Show more", "Add to Mendeley", "Share", and "Cite", along with the DOI link <https://doi.org/10.1016/j.margeo.2013.03.006> and a "Get rights and content" link.



The screenshot shows the "WEB OF KNOWLEDGE" interface, specifically the "THE DATA CITATION INDEX" section. The page features a map of the world with a network of blue lines connecting various locations, representing the global reach of the citation index. A large orange circle in the center of the map contains the text "THE DATA CITATION INDEX ON WEB OF KNOWLEDGE". To the left of the map, there is a "What is it?" button with a video icon. The top navigation bar includes links for "ABOUT", "PRODUCTS & TOOLS", "BENEFITS & RESOURCES", "TRAINING & SUPPORT", "NEWS & EVENTS", "CONTACT US", and "Site Search". The search bar is located at the top right.

Data Management Planning

- Concept/Proposal Development
 - Are existing data available?
- Data Acquisition Plan
 - Sensor Calibration
 - Survey Plans
 - Data Analysis + Reduction
 - Data Documentation
- Data Management Plan (DMP)
 - How will you preserve & document data?
 - What are submission requirements?

U.S. National Science Foundation

Find Funding ▾ How to Apply ▾ Manage Your Award ▾ Focus Areas ▾

Home / Policies / Policies and Procedures / Division of Ocean Sciences (OCE) Sample and Data Policy

Policies and Procedures

Division of Ocean Sciences (OCE) Sample and Data Policy

All NSF proposals must include a Data Management and Sharing Plan that describes what data/samples will be collected, what analyses will be done, and how the project will provide open and timely access to meta data, data, preserved or archived samples, derived products (e.g., models and model output), and other information on the project during and after the project's completion. The Data Management and Sharing Plan also must specifically discuss how the investigators will achieve the specific OCE data archiving and reporting requirements described below in this document. If the project is not expected to generate new data, samples or derived data products, the Data Management and Sharing Plan can include a statement that no detailed plan is needed, accompanied by a clear justification. See the [NSF Proposal & Award Policies & Procedures Guide \(PAPPG\)](#), Chapter II.D.2 for additional information.

NSF Principal Investigators (PIs) are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections, meta data, and other supporting materials created or gathered in the course of work under NSF grants.

OCE Data and Sample Repositories

Division of Ocean Sciences (GEO/OCE)

Programs and Funding Opportunities +
Career Opportunities +
About +
News and Events +
Resources -

Become a GEO Reviewer
Career Development Resources
Data Management and Sharing Plan Guidance
OCE Data and Sample Repositories
Policies for Resubmitting Declined Proposals
OCE Forms and Resources for Research at Sea
OCE Environmental Compliance Documents
Biological Oceanography Program Additional Information

Refer to the U.S. National Science Foundation (OCE) Sample and Data Policy (NSF 24-124) data management and sharing plans.

On this page

- Data repositories
- Sample repositories

Data repositories

Biological and Chemical Oceanography Data Office (BCO-DMO) ↗
Research projects funded by the NSF Biological Oceanography Program must be registered with the BCO-DMO. Data deposited directly with BCO-DMO or archived elsewhere and data link as well as relevant metadata should be provided to BCO-DMO.

CLIVAR and Carbon Hydrographic Data Office (CHDO) ↗
CHDO supports oceanographic research by providing access to high quality, global, vessel-based conductivity, temperature and depth and hydrographic data from the Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP), World Ocean Circulation Experiment (WOCE), Climate and Ocean - Variability, Predictability and Change (CLIVAR) and other repeat hydrography programs.

Marine Geoscience Data System (MGDS) ↗
MGDS is a data repository that provides free public access to a curated collection of marine geophysical data products and complementary data collected throughout the global oceans.

National Deep Submergence Facility (NDSF) ↗
The National Deep Submergence Facility (NDSF) manages data collected by the submersible Alvin, ROV Jason and AUV Sentry.

Division of Ocean Sciences (OCE) Sample and Data Policy

All NSF proposals must include a Data Management and Sharing Plan that describes what data/samples will be collected, what analyses will be done, and how the project will provide open and timely access to meta data, data, preserved or archived samples, derived products (e.g., models and model output), and other information on the project during and after the project's completion. The Data Management and Sharing Plan also must specifically discuss how the investigators will achieve the specific OCE data archiving and reporting requirements described below in this document. If the project is not expected to generate new data, samples or derived data products, the Data Management and Sharing Plan can include a statement that no detailed plan is needed, accompanied by a clear justification. See the [NSF Proposal & Award Policies & Procedures Guide \(PAPPG\)](#), Chapter II.D.2 for additional information.

NSF Principal Investigators (PIs) are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections, meta data, and other supporting materials created or gathered in the course of work under NSF grants.

Data and sample archiving requirements

The Division of Ocean Sciences requires that meta data files, full data sets, derived products and physical collections must be made publicly accessible upon publication, or within two (2) years of collection, whichever comes first. This includes software and derived data products (e.g., model results, output, and workflows). A brief description of [preferred data and physical collection archives and centers](#) and their criteria for submission can be found on the [OCE website](#) or through contact with the cognizant Program Officer of the given award. Data should be in a format that is easily accessible; data output in a format that is not readily accessible should be accompanied by a readme file and a script to allow reading it. Any limit on access to original data, samples, or other information beyond the two-year moratorium period must be based on compelling justification, documented in the Data Management and Sharing Plan of the proposal, or approved by the cognizant Program Officer. Such exceptions have been, and are likely to remain, rare.



Field Data

- Facility-Managed Data
 - Ship (R2R)
 - Raw sensor data
 - Vehicle Data (WHOI + MGDS)
 - Raw sensor data
 - First-order at-sea products
- PI-Managed Data
 - Documentation
 - Cruise report
 - Sample metadata
 - Processing metadata
 - Physical samples
 - Science party instrumentation
 - Derived data products



Sealog

AT42-22 > J0-1245 turning on ICL2

FREEFORM turning on ICL2

guret at 2020-02-02T01:40:55.814Z

SciCam BrowCam PixCam AftCam

Navigation

Time	Latitude	Longitude	Depth	Altitude	Heading	Pitch	Roll	Original Net Data
2020-02-02T01:40:55.814Z	18.54621163	-41.71523232	-4959.399	13.0	45.72	-14.7	3	18.54621163 deg
								41.71523232 deg
								45.68 deg
								4953.30 meters
								7.09 meters

2020-02-02T01:40:55.814Z 2020-02-02T01:40:55.814Z

Filtered Events

2020-02-02T01:40:55.814Z - querying: "turning on ICL2"

2020-02-02T01:53:47.482Z - querying: "sample type: 'IGT Fluid', taking J1245/IGT3 at Beebe Woods max temperature 348 C"

2020-02-02T01:55:13.707Z - querying: "temperature very stable at 348 C"

2020-02-02T01:56:14.087Z - querying: "holding IGT3"

2020-02-02T01:57:43.562Z - querying: "heading for lung snack"

2020-02-02T02:01:01.894Z - querying: "lung snack found"

2020-02-02T02:01:23.460Z - querying: "landed at lung snack preparing to sample starting with IGT1"

2020-02-02T02:05:21.116Z - querying: "sample type: 'IGT Fluid', taking J1245/IGT1 at Lung Snack max temperature 280 C"

2020-02-02T02:08:21.802Z - querying: "stowing IGT"

2020-02-02T02:08:28.359Z - querying: "stowing IGT"

Event Filter

Search

e.g. fish

Author

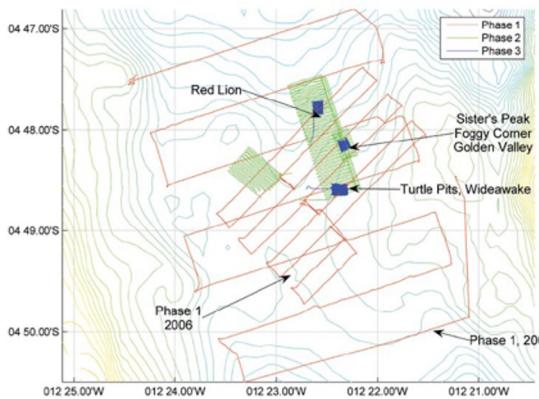
e.g. jsmith

Start Date/Time (UTC)

Stop Date/Time (UTC)

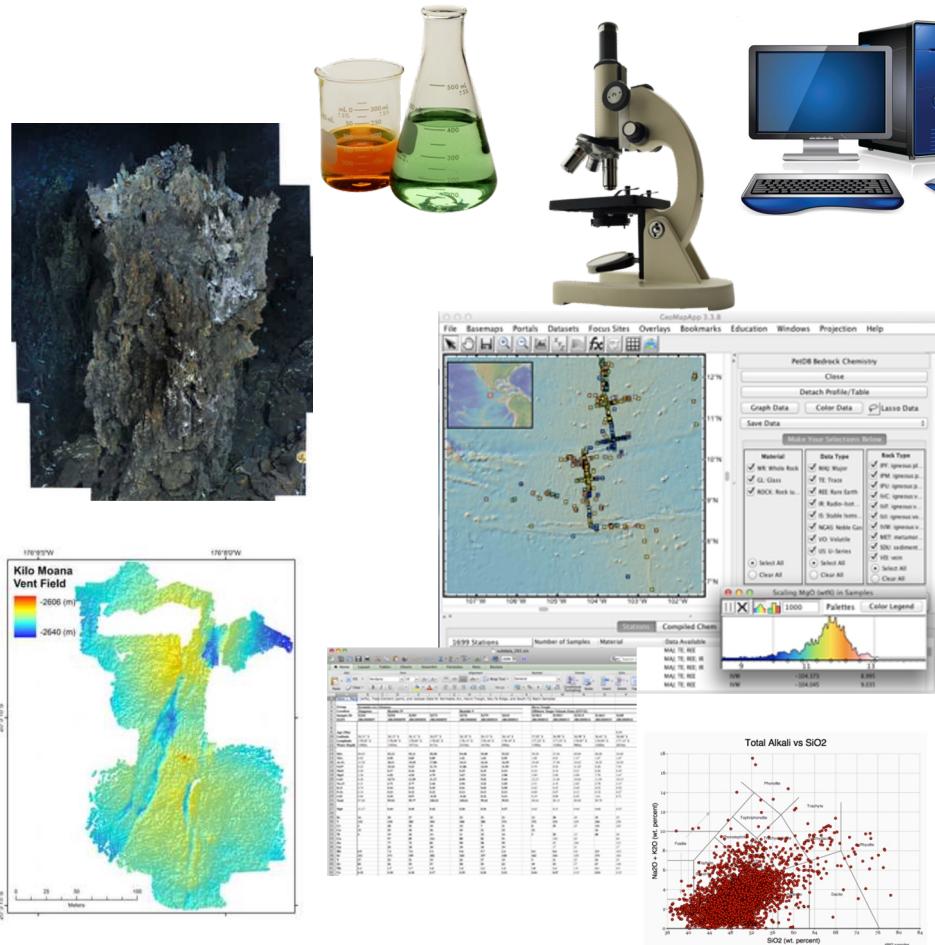
Reset Filter

Sealog is licensed under the MIT public license



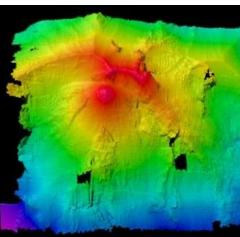
Processed/Derived Data

- Which data should be preserved?
 - Data supporting publications
 - Processed data of value
 - Results of lab analysis
- Where should it be curated?
 - Domain-specific repository?
 - What are requirements of repository?
- Documentation
 - What does a new user need to know?
 - How were products generated?
 - What are caveats of data?



Deep-Submergence Data Resources

- WHOI/NDSF (ndsf.whoi.edu)
 - Dive metadata & data at WHOI Data Library
 - NDSF vehicles
- Marine Geoscience Data System (marine-geo.org)
 - Dive metadata, field & derived data
 - Navigation, geophysical data, event logs, bottom photos
 - Data from many operators/vehicles
 - SOI: ROV SuBastian
 - NDSF: Alvin, Jason, Sentry
 - Nautilus: Hercules
 - MBARI Mapping AUV
 - REMUS
 - LEGACY: DSL120, IMI130, ABE
- GMRT (www.gmrt.org)
 - Bathymetry grids/images
- NOAA/NCEI (www.ncei.noaa.gov)
 - OER dive metadata & data
 - Searchable dive video archive
- Scientific publications
- Other...



Sealog
AT422 > J0-1245 > Replay
FREE FORM turning on ICL2
Navigation
Time: 2020-02-02T01:40:55.814Z Depth: -899.399 meters Latitude: 18.14035454 0deg
Latitude: -81.718232 0deg
Longitude: 335.76 0deg Heading: 147.0 deg
Longitude: 499.3 0deg
Depth: 975.76 meters Roll: 3.3 deg
Altitude: 75 meters
Original Nav Data
Time: 2020-02-02T01:40:55.814Z Depth: -899.399 meters Latitude: 18.14035454 0deg
Latitude: -81.718232 0deg
Longitude: 335.76 0deg Heading: 147.0 deg
Longitude: 499.3 0deg
Depth: 975.76 meters Roll: 3.3 deg
Altitude: 75 meters
Event Filter
Search
Author
Start Date/Time (UTC)
Stop Date/Time (UTC)
Reset Filter
Sealog is licensed under the MIT public license

NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
formerly the National Oceanographic Data Center (NODC) [more on NCEI](http://www.ncei.noaa.gov)

Home Access Data Submit Data Public Outreach About

You are here: [Home](#) > OER Video Portal

OER Video Portal Feedback | Help

Oceans Exploration and Research

View Basket

OER Video Portal

Search, discover and access video data from OER sponsored missions on NOAA Ship Okeanos Explorer and other vessels. Please use this form to share your feedback on the OER Video Portal.

Keywords

Apply OR AND to multiple keywords (separate with commas)
Match keywords anywhere exactly
To choose from a specific dive NN use DIVENN.

Observation Dates

Start date: End date: Autocomplete

Depth Range

Min. Depth (in meters): Max. Depth (in meters):

Video Type

Segment Street Highlight

Camera Location (by Segments)

Camera Platform ROV Other

Multiple selections are allowed. Use Click/Shift/Click for lists, Click/Control/Shift/Control for segments. To select multiple segments, use Ctrl/Shift/Alt/Shift/Alt.

Geographic Coverage

Hold the "shift" key and drag to select an area.

Intersecting Fully within Anywhere

North: 90 West: 180 Global East: 180

South: 90

File Name: COL.20111020.081729042.0046514.tif
Date: 2011-10-20 08:17:29
Latitude: 9.50693846
Longitude: 104.2428492
Download Full Resolution Image
View Expedition Data Sets

MGDS MARINE GEOSCIENCE DATA SYSTEM

File Name: COL.20111020.081729042.0046514.tif
Date: 2011-10-20 08:17:29
Latitude: 9.50693846
Longitude: 104.2428492
Download Full Resolution Image
View Expedition Data Sets

Overlays
Data Set(s)
Photo(s)

Google

20 km

Terms of Use



Navigating Data Management

- Plan ahead
- Know what resources are available
 - Software Tools
 - Guidelines & Templates
- Communicate
 - Upstream (Operations Team)
 - Downstream (Data Managers)
- Organize consistently
- Document contemporaneously
- Data is a valuable community resource
- Participate!
 - Metadata & data format standards
 - System usability





Which Repository?

- Know data policies
- Seek **domain-specific** repositories
- System features to consider
 - Long-term Archiving
 - Data Usage Reports
 - Data Publication
 - User Support
 - Usability
 - Interoperability

*Reach out to data professionals -
we're here to help!!*



PetDB



Questions?

ferrini@ldeo.columbia.edu