

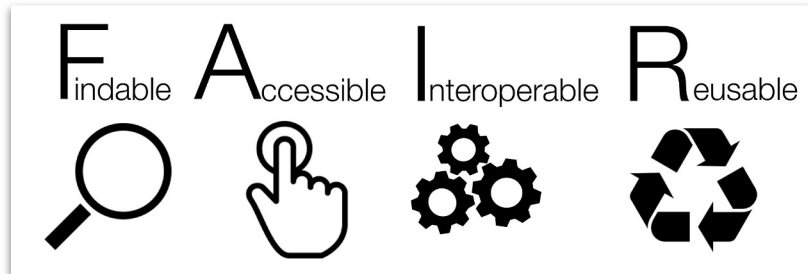
DeSSC New-User Program: Data Management Overview

Vicki Ferrini, LDEO

Open Data Access

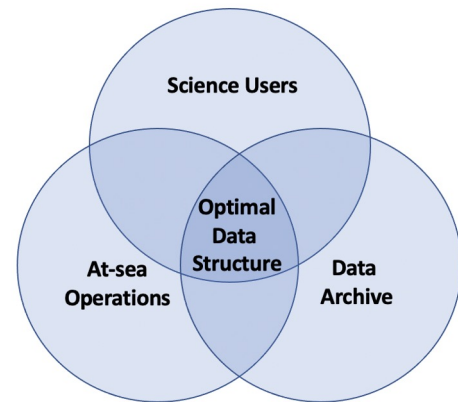
- Driven by:
 - Federal and funding-agency requirements
 - Journal and publisher expectations
 - High acquisition costs that incentivize reuse
- Enables New Opportunities:
 - Tracking spatial & temporal change
 - Scientific transparency and reproducibility
 - Cross-disciplinary data synthesis
 - New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

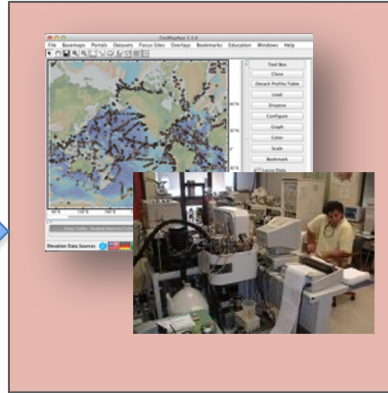


Accessibility is a shared responsibility — and shared benefit.

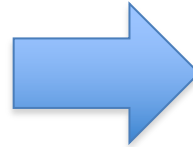
Historic Scientific Workflow



Data Acquisition



Data Processing
& Interpretation



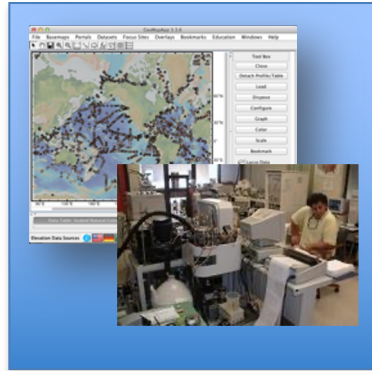
Publications



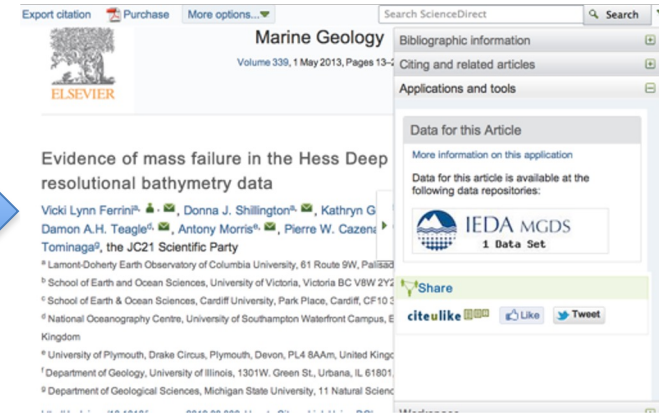
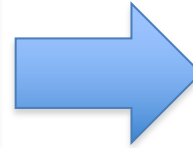
Modern Scientific Workflow



Data Acquisition



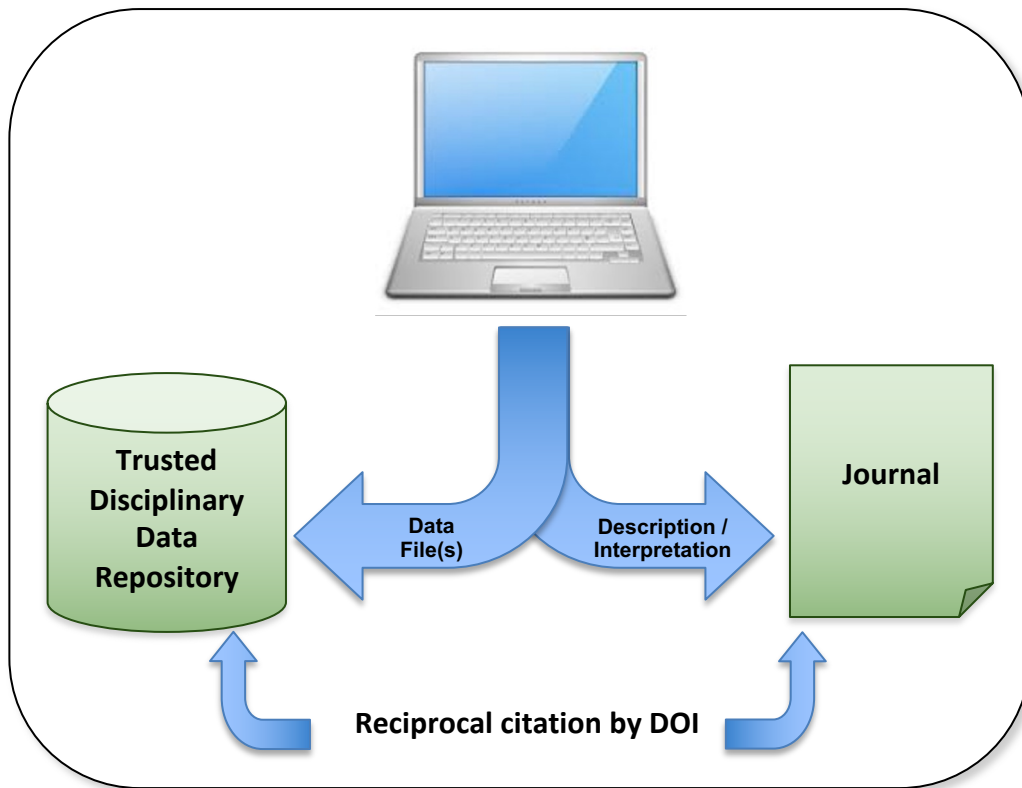
Data Processing
& Interpretation



Publications



“Best Practice”



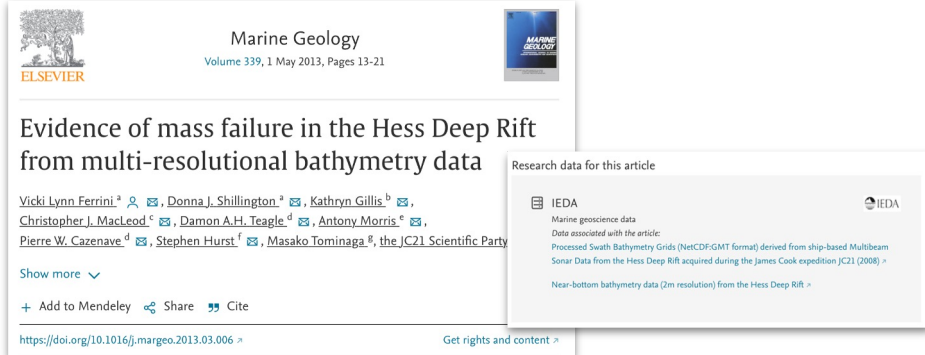
What's in it for you?

- 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 the top of a research article from the journal *Marine Geology*, Volume 339, 1 May 2013, Pages 13-21. The article title is "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, Masako Tominaga, and the JC21 Scientific Party. A sidebar on the right titled "Research data for this article" lists data from IEDA: "Marine geoscience data", "Data associated with the article: 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)", and "Near-bottom bathymetry data (2m resolution) from the Hess Deep Rift".



The screenshot shows the "WEB OF KNOWLEDGE" logo at the top, with a navigation bar including "ABOUT", "PRODUCTS & TOOLS", "BENEFITS & RESOURCES", "TRAINING & SUPPORT", "NEWS & EVENTS", and "CONTACT US". Below the navigation bar, the main heading is "THE DATA CITATION INDEX™" with the tagline "CONNECTING THE DATA TO THE RESEARCH IT INFORMS". A central graphic features a globe with various icons (buildings, people, etc.) connected by lines to a central orange circle labeled "THE DATA CITATION INDEX™ ON WEB OF KNOWLEDGE™". A "What is it?" button is visible on the left.

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 / Directorate for Geosciences (GEO) / Division of Ocean Sciences (GEO/OCE) / OCE Data and Sample Repositories

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 (BCO-DMO) [↗](#)

Research projects funded by the NSF Biological Oceanography programs 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 (CCHDO) [↗](#)

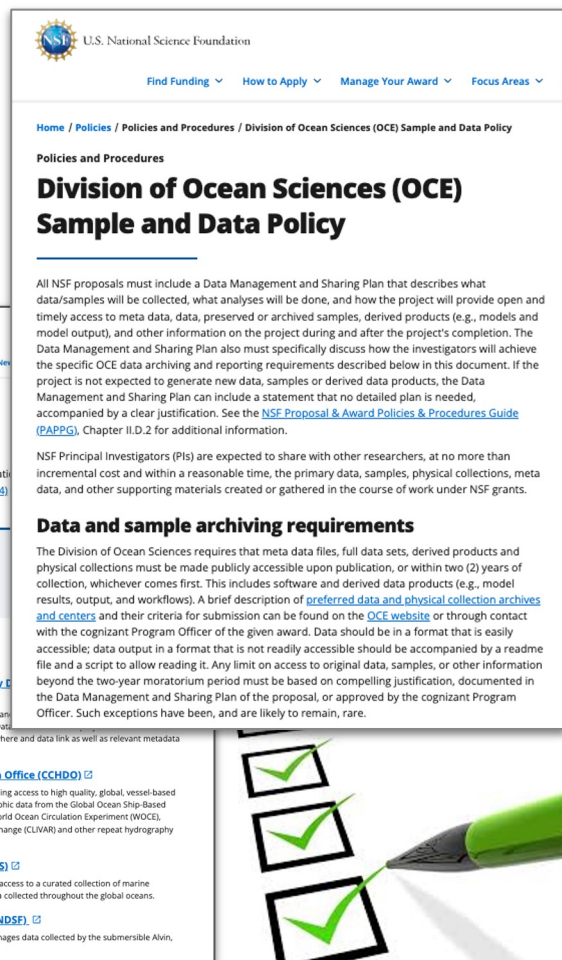
CCHDO supports oceanographic research by providing access to high quality, global, vessel-based conductivity, temperature and depth and hydrographic data from the Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP), World Ocean Circulation Experiment (WOCE), Climate and Ocean - Variability, 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.



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

Division of Ocean Sciences (OCE) Sample and Data Policy


Policies and Procedures

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 \(PAPPs\)](#), 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.

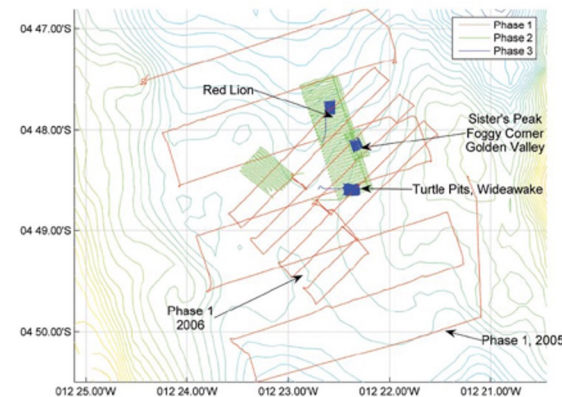
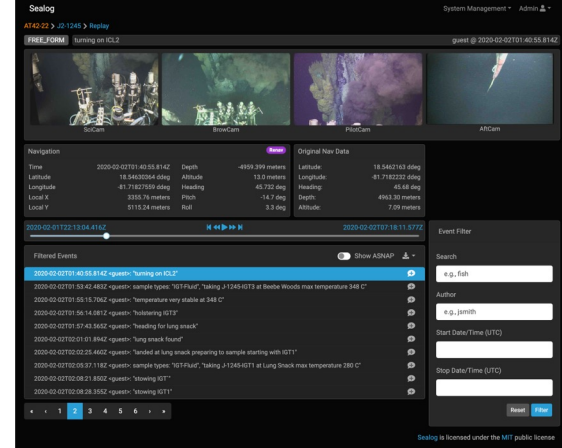
Data and sample archiving requirements

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.




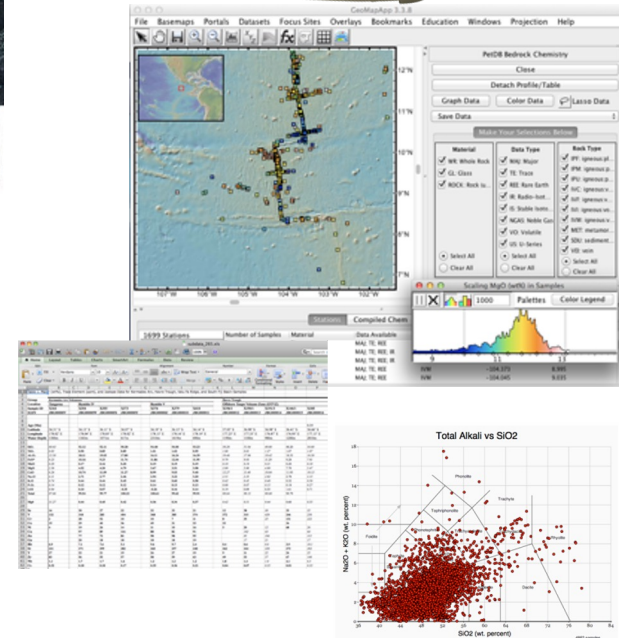
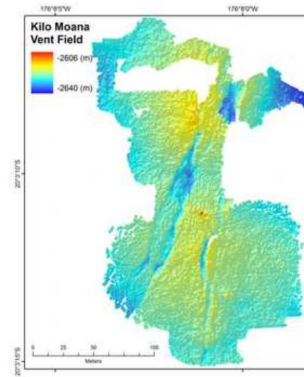
Field Data

- 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



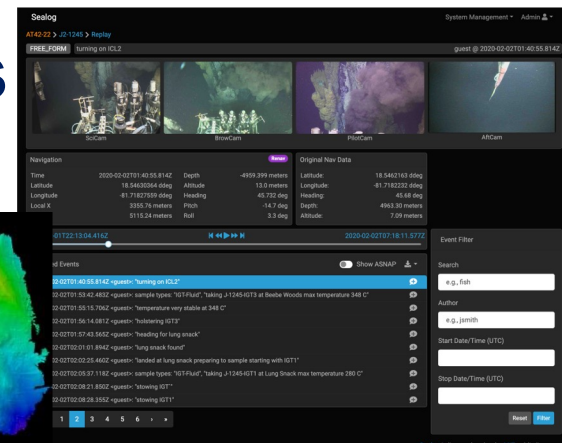
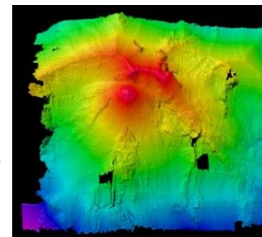
A collage of four images representing different scientific fields: chemistry (beakers with liquids), biology (a microscope), and computer science (a desktop computer).

- 



Deep-Submergence Data Resources

- WHOI/NDSF (ndsf.who.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...





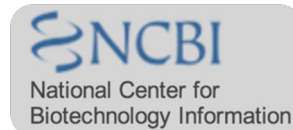

- [illegible]



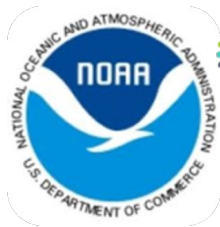
Which Repository?

- 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