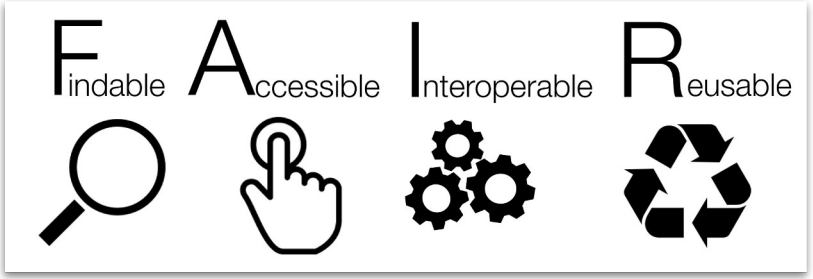
DeSSC New-User Program: Data Management Overview

Vicki Ferrini, LDEO

Lamont-Doherty Earth Observatory
Columbia Climate School

Open Data Access

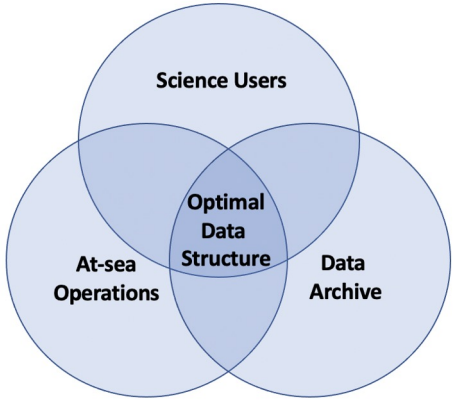


- Driven by:
  - Federal and funding-agency requirements
  - Journal and publisher expectations
  - High acquisition costs that incentivize reuse
- Enables New Opportunities:
  - Tracking spatial & temporal change
  - Scientific transparency and reproducibility
  - Cross-disciplinary data synthesis
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

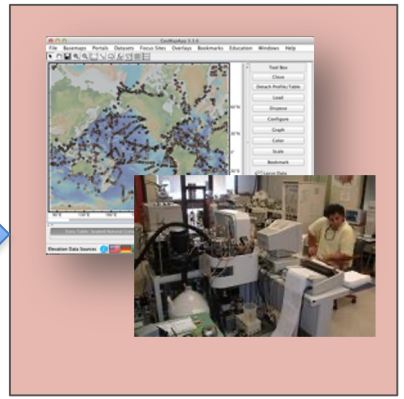
- **Data producers** → share, document, and preserve
- **Data stewards** → curate, maintain, and support discovery
- **Data users** → cite, credit, and contribute improvements
- **Journals & agencies** → reinforce best practices



Accessibility is a shared responsibility — and shared benefit.

Historic Scientific Workflow

Data Acquisition



Data Processing
& Interpretation

PRIVATE
PROPERTY
NO PUBLIC
ACCESS

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

[CH]

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, USA*
[b] *Lamont-Doherty Geological Observatory of Columbia University, Palisades, NY 10964, USA*
[c] *Department of Geological Sciences, Columbia University, Palisades, NY 10964, USA*

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

ABSTRACT

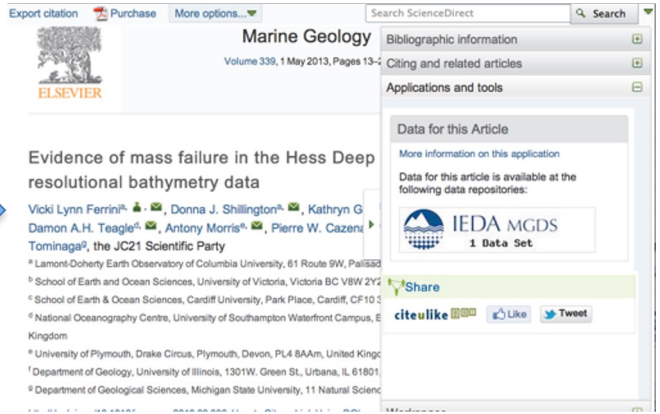
Using the near-bottom ARGO imaging system, we visually and acoustically surveyed the narrow (< 200 m wide) axial zone of the fast-spreading East Pacific Rise (EPR) along 83 km of its length (9°09′–54′N), discovered the Venture Hydrothermal Fields, and systematically mapped the distribution of hundreds of hydrothermal features relative to other fine-scale volcanic and tectonic features of the ridge crest. The survey encompasses most of a 2nd order ridge segment and includes at least ten 4th order (5–15 km) segments defined by bends or small lateral offsets of the ridge crest or axis (Desvals). 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 and by changes in the characteristics of the axial zone (axial lava age, extent of fissuring, axial morphology and structure, etc.) across segment boundaries. The distribution and along-strike variability of hydrothermal features corresponds closely to the morphotectonic/structural segmentation of the ridge. On the 2nd order scale, we find that high *T* hydrothermal activity correlates with: (1) shallowing of the axial magma chamber (AMC) reflector to depths < 1.7 km beneath 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 summit graben (ASG); however, the extent of volcanic collapse along the ASG revealed by the ARGO survey adds to evidence that on fast-spreading ridges it is an elongate volcanic caldera rather than a tectonic 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 20 m of the margins of the ASC. Despite the significant extent of our track coverage outside 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 ASC is very narrow (40–70 m), (2) the AMC reflector is most shallow (< 1.55 km beneath the axial

Publications

Modern Scientific Workflow

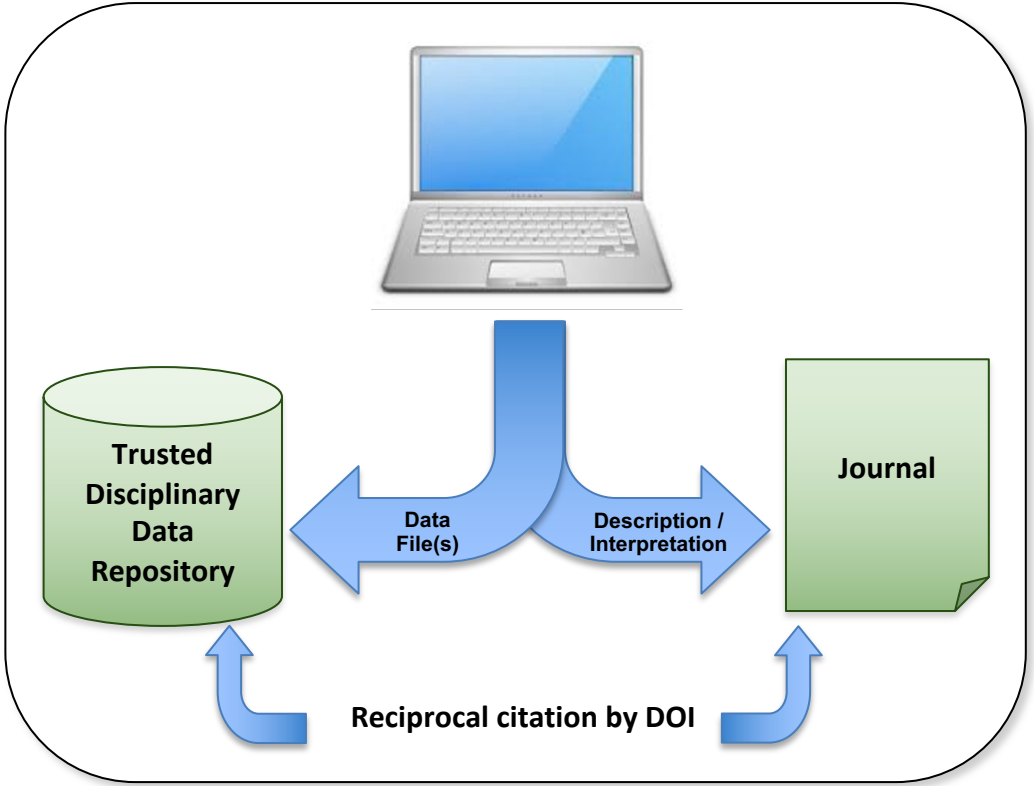
Data Acquisition

Data Processing
& Interpretation



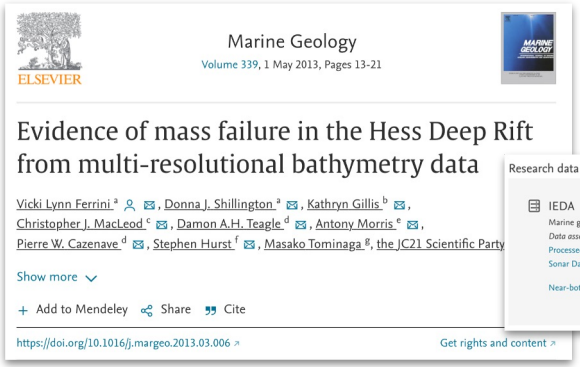
Publications

“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



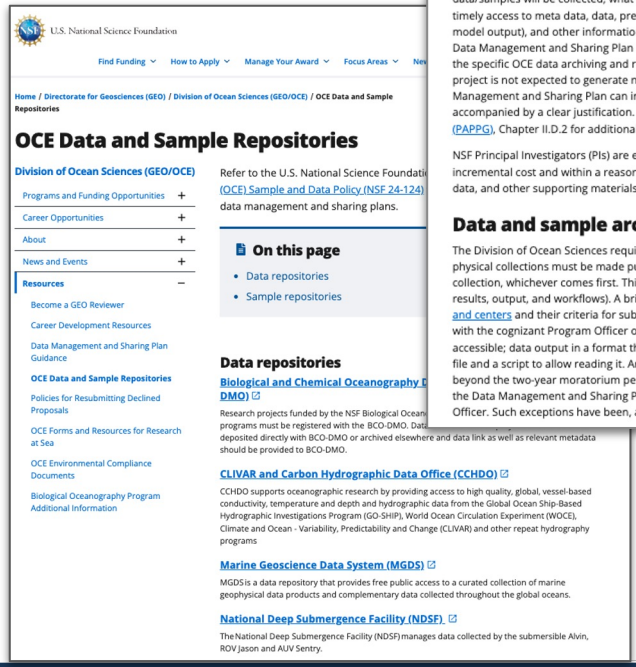
"The coolest thing to do with your data will be thought of by someone else."

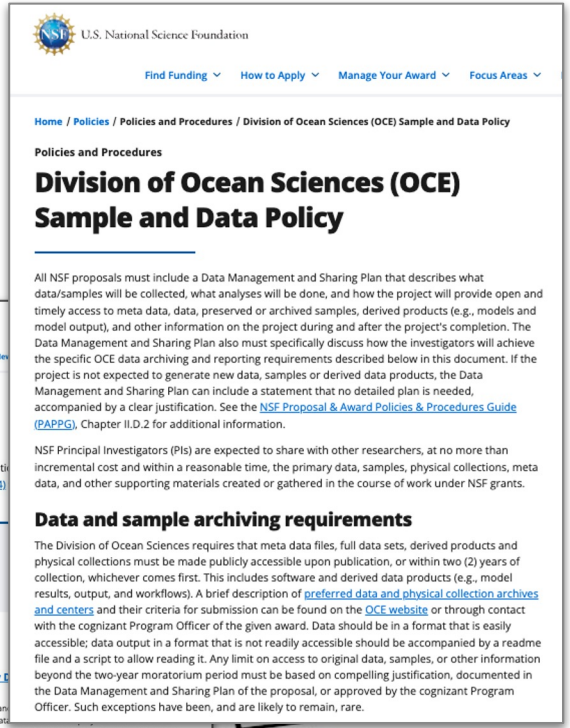
Rufus Pollock

Cambridge University and Open Knowledge Foundation

Data Management Planning

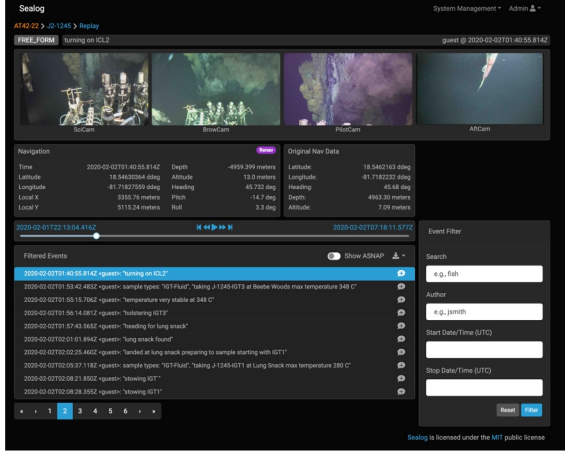
- Concept/Proposal Development
  - Are existing data available?
- Data Acquisition Plan
  - Sensor Calibration
  - Survey Plans
  - Data Analysis + Reduction
  - Data Documentation
- Data Management Plan (DMP)
  - How will you preserve & document data?
  - What are submission requirements?

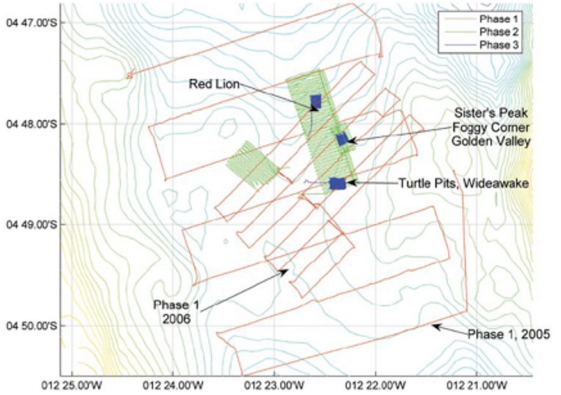




Field Data

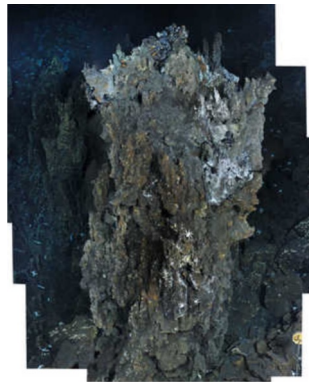
- Facility-Managed Data
  - Ship (R2R)
    - Raw sensor data
  - Vehicle Data (WHOI + MGDS)
    - Raw sensor data
    - First-order at-sea products
- PI-Managed Data
  - Documentation
  - Cruise report
  - Sample metadata
  - Processing metadata
  - Physical samples
  - Science party instrumentation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

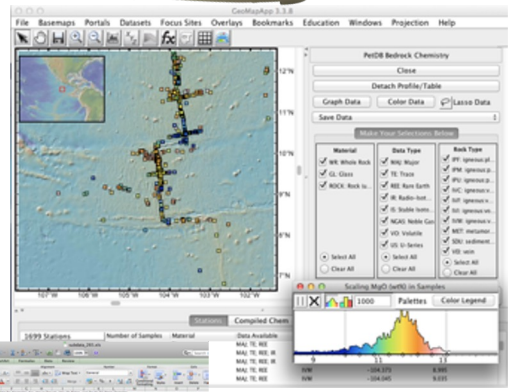


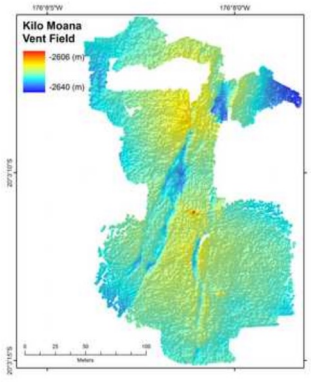


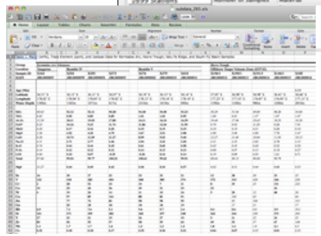
Processed/Derived Data

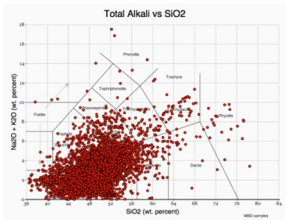
- Which data should be preserved?
  - Data supporting publications
  - Processed data of value
  - Results of lab analysis
- Where should it be curated?
  - Domain-specific repository?
  - What are requirements of repository?
- Documentation
  - What does a new user need to know?
  - How were products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





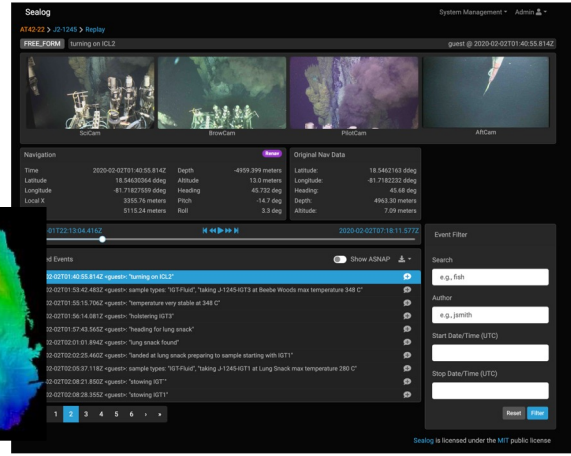


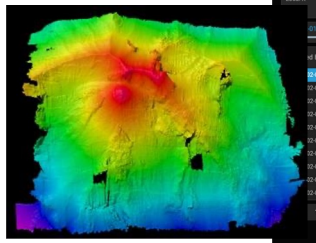


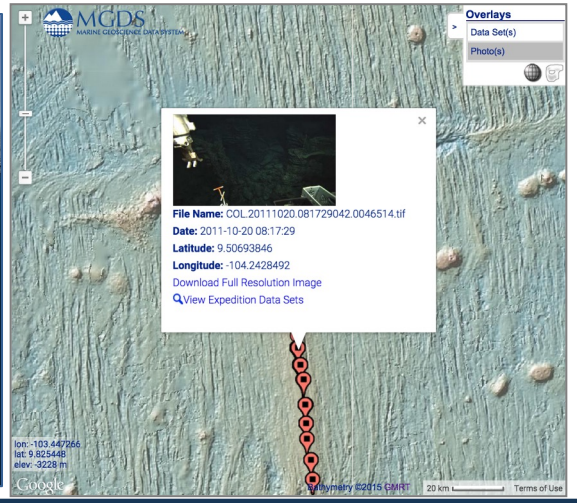


Deep-Submergence Data Resources

- 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…







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!!**

Questions?

ferrini@ldeo.columbia.edu