

# **DESSC New-User Workshop:** *Data Management Overview*

Vicki Ferrini

*Lamont-Doherty Earth Observatory*



# Increasing Emphasis on *Open Data Access*

- Federal Data Policies
- Acquisition Costs & Data Volumes
- Spatial & Temporal Change
- Scientific Reproducibility
- New Possibilities for Analysis
- Big Data

## Division of Ocean Sciences Sample and Data Policy



National Science Foundation

May 2011



SEARCH

NSF Web Site

[HOME](#) | [FUNDING](#) | [AWARDS](#) | [DISCOVERIES](#) | [NEWS](#) | [PUBLICATIONS](#) | [STATISTICS](#) | [ABOUT](#) | [FastLane](#)

[Find Funding](#) • [Recent Funding](#) • [Due Dates](#) • [Preparing Proposals](#) • [About Funding](#)

Cloud Computing Opens  
New Possibilities →

FEATURES 1 2 3 4 ▶

## News

### White House issues directive supporting public access to publicly funded research

Timothy Vollmer, February 22nd, 2013



Seal of the United States Office of Science  
and Technology Policy / Public Domain

Today, the White House issued a Directive supporting public access to publicly-funded research.

John Holdren, Director of the Office of Science and Technology Policy, "has directed Federal agencies with more than \$100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research."

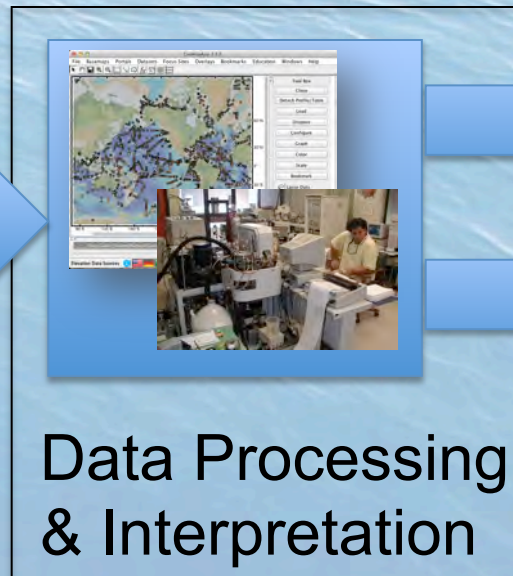
Each agency covered by the Directive (54 KB PDF) must "Ensure that the public can read, download, and analyze in digital form final peer reviewed manuscripts or final published documents within a timeframe that is



# Historic Scientific Workflow

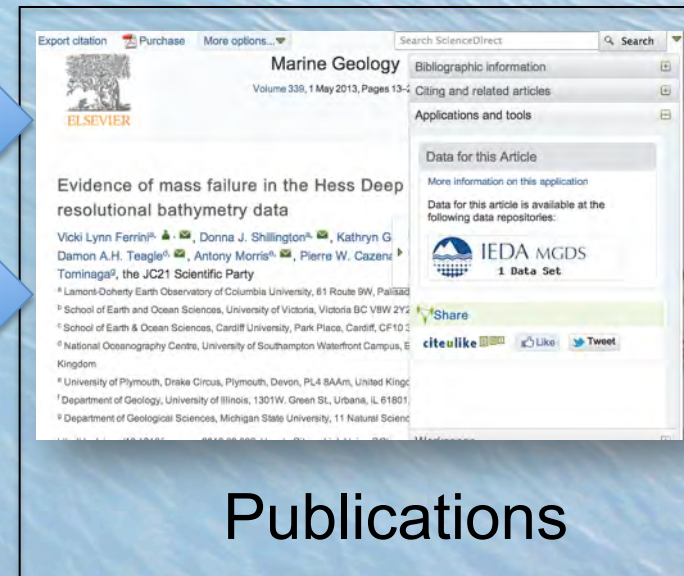


Data Acquisition



Data Processing  
& Interpretation

**PRIVATE  
PROPERTY  
NO PUBLIC  
ACCESS**



Publications





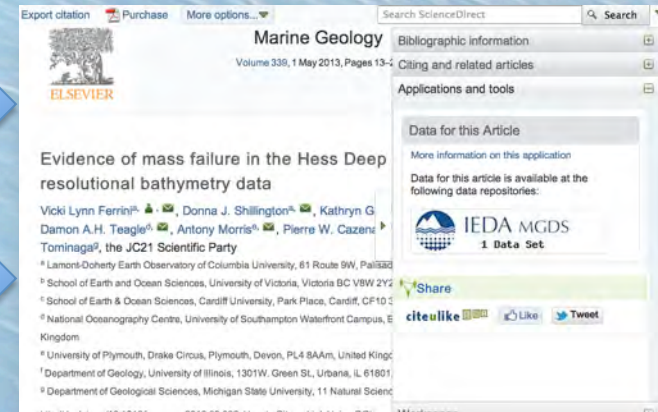
# Modern Scientific Workflow



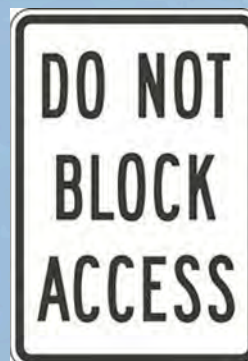
Data Acquisition



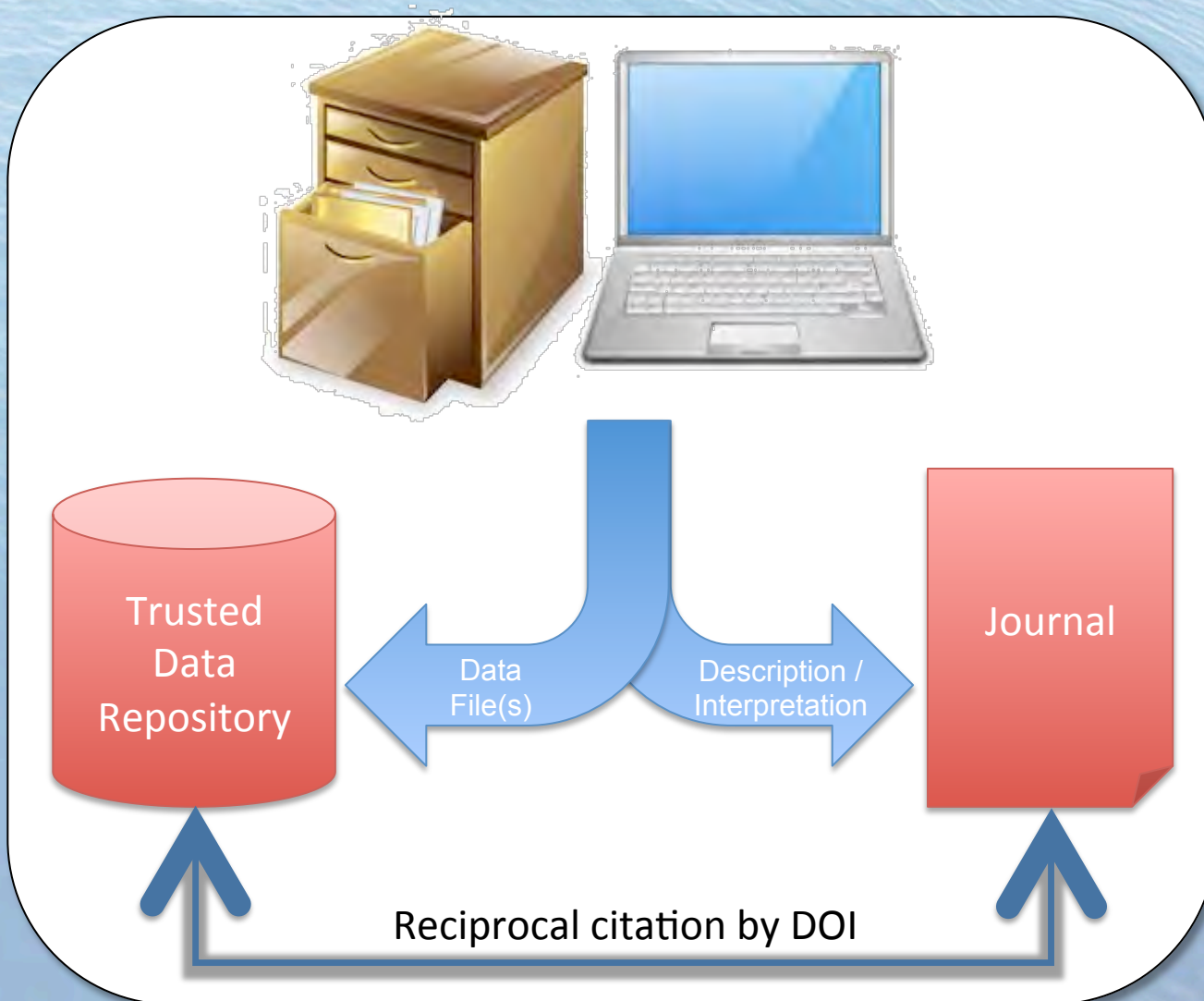
Data Processing  
& Interpretation



Publications



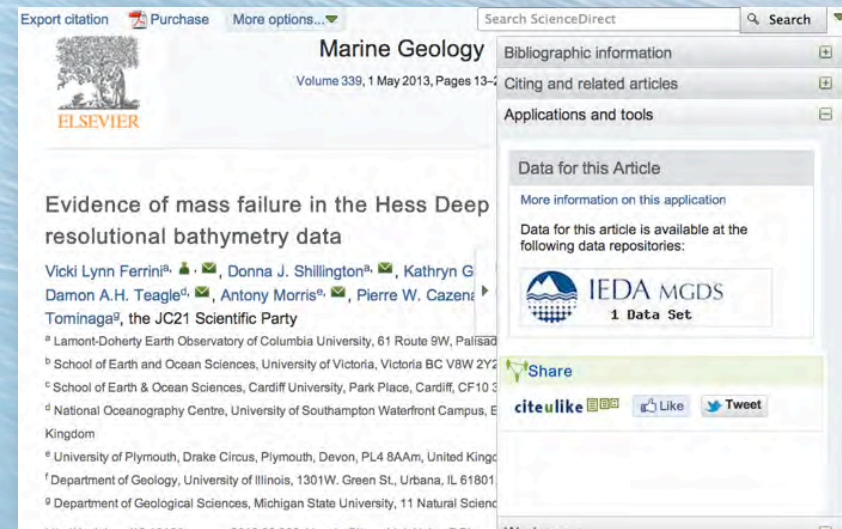
# “Best Practice”





# What's in it for you?

- Scientific Integrity & Reproducibility
- Collaborative opportunities
- Attribution (Citation)
- Increase the impact of your research
- Preserve data for your own future use
- Compliance with Data Policies
- Education & Outreach



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

*Rufus Pollock*

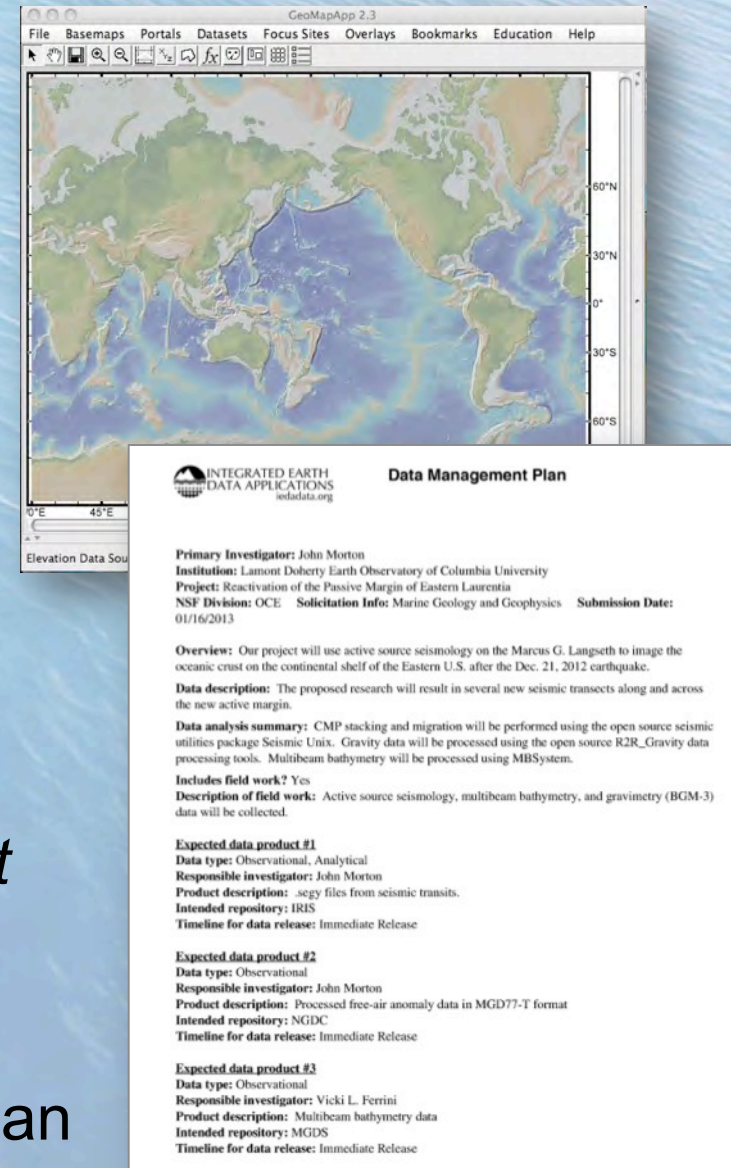
*Cambridge University and Open Knowledge Foundation*



# Data Management Planning

- Concept/Proposal Development
  - *Are Existing Data Available?*
- Data Acquisition Plan
  - Sensor Calibration
  - Survey Plan
  - Data Analysis + Reduction
  - Data Documentation
- Data Management Plan (DMP)
  - Required in NSF Proposals
  - *How will you preserve & document your data?*

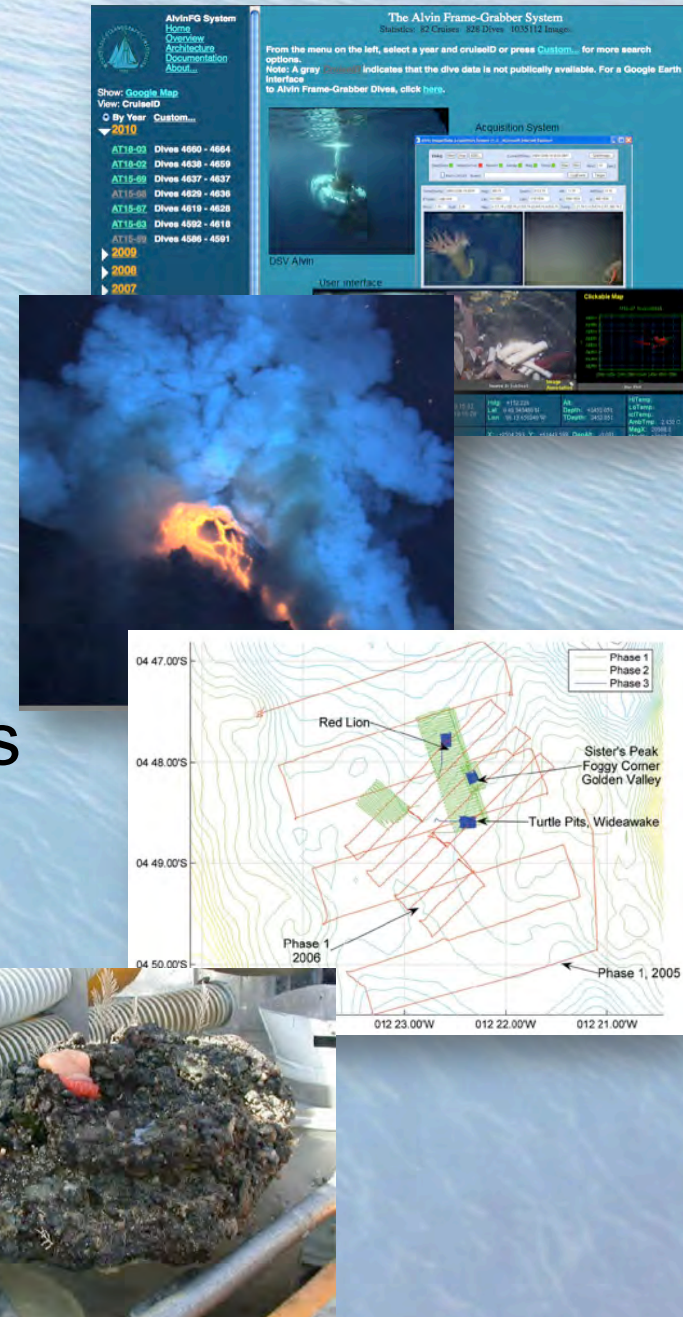
<http://www.iedadata.org/compliance/plan>





# Field Data

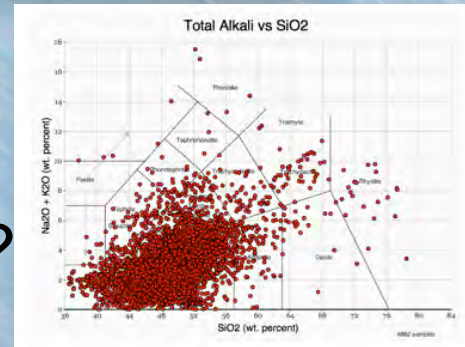
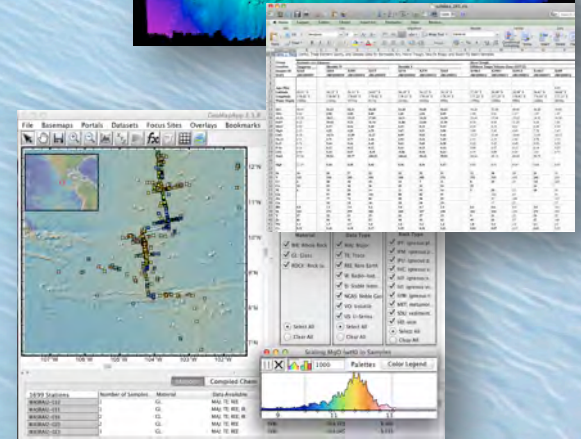
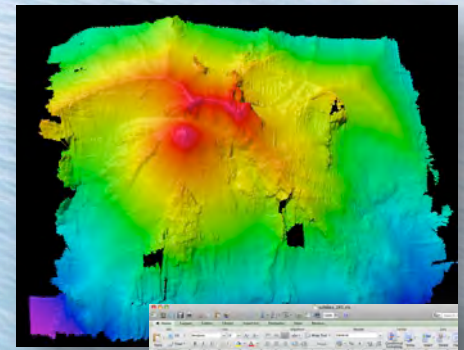
- Facility-Managed Data ✓
  - Ship (R2R)
    - Raw sensor data
  - Vehicle Data (NDSF)
    - Raw sensor data
    - First-order at-sea products
- PI-Managed Data
  - Documentation
  - Cruise report
  - Sample metadata
  - Processing metadata
  - Physical samples
  - Science party instrumentation





# Derived Data

- 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?*





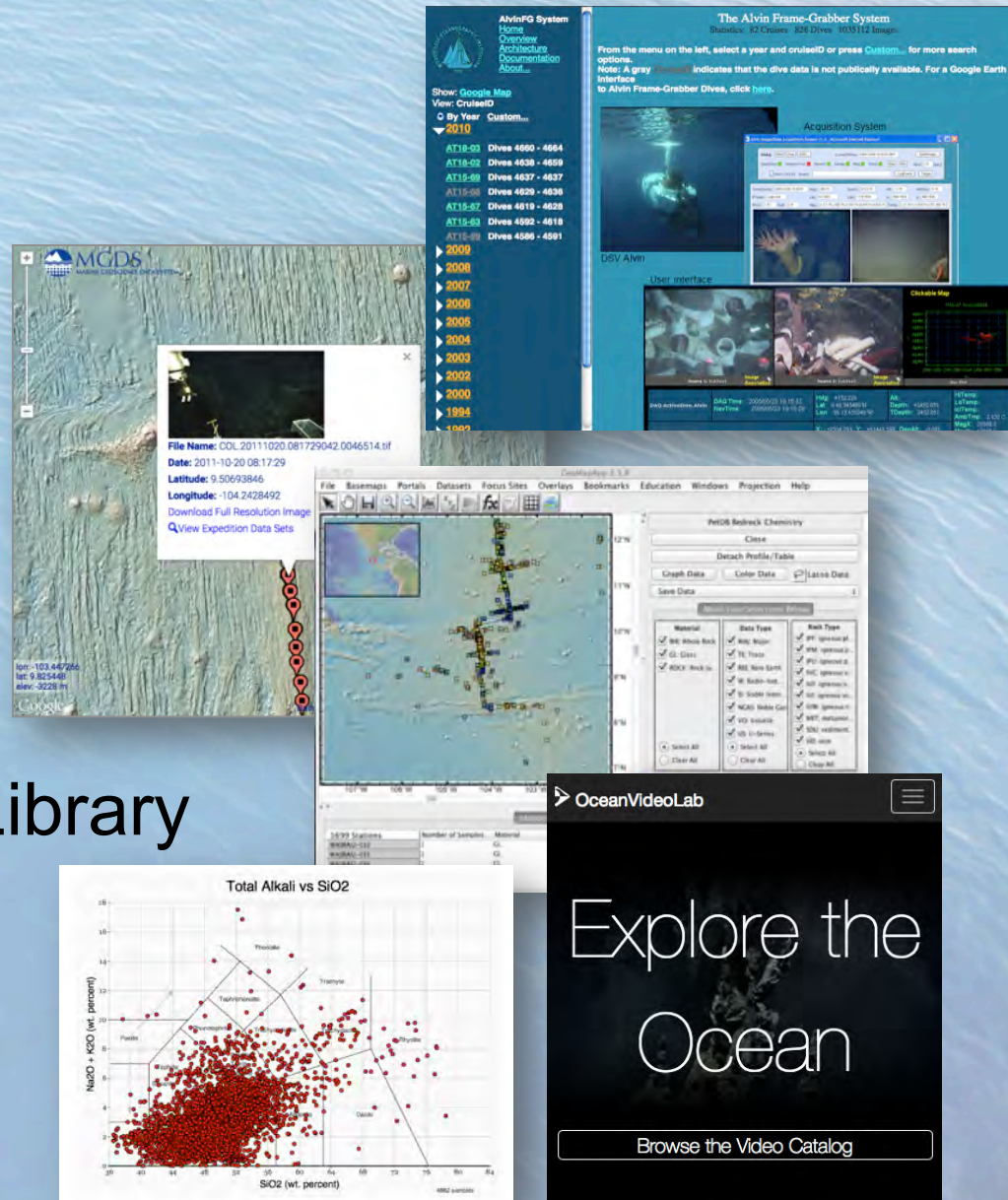


-



# NDSF-Related Data Resources

- [4dgeo.who.edu](http://4dgeo.who.edu)
  - Jason Virtual Van
  - Alvin FrameGrabber
- [iedadata.org](http://iedadata.org)
  - EarthChem Library
  - EarthChem Portal
  - GeoMapApp
  - GMRT
  - Marine-Geo Digital Library
  - PetDB
  - SESAR
- [oceanvideolab.org](http://oceanvideolab.org)

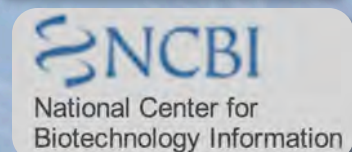
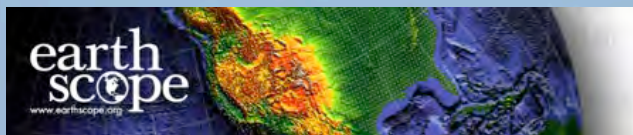






# Which Repository?

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





# Questions?

[ferrini@Ideo.columbia.edu](mailto:ferrini@Ideo.columbia.edu)