

DESSC New-User Program: *Data Management Overview*

Vicki Ferrini

Lamont-Doherty Earth Observatory



Open Data Access

Driven by:

- Funding Agency Requirements
- Journal Requirements
- Acquisition Costs & Data Volumes

Enables New Opportunities:

- Spatial & temporal change
- Scientific reproducibility
- New possibilities for analysis



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

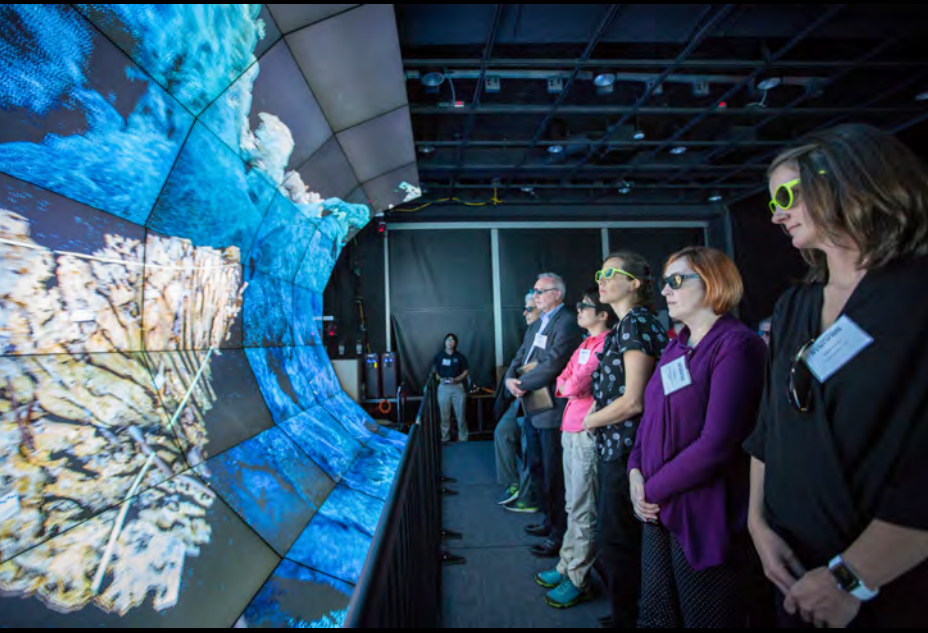
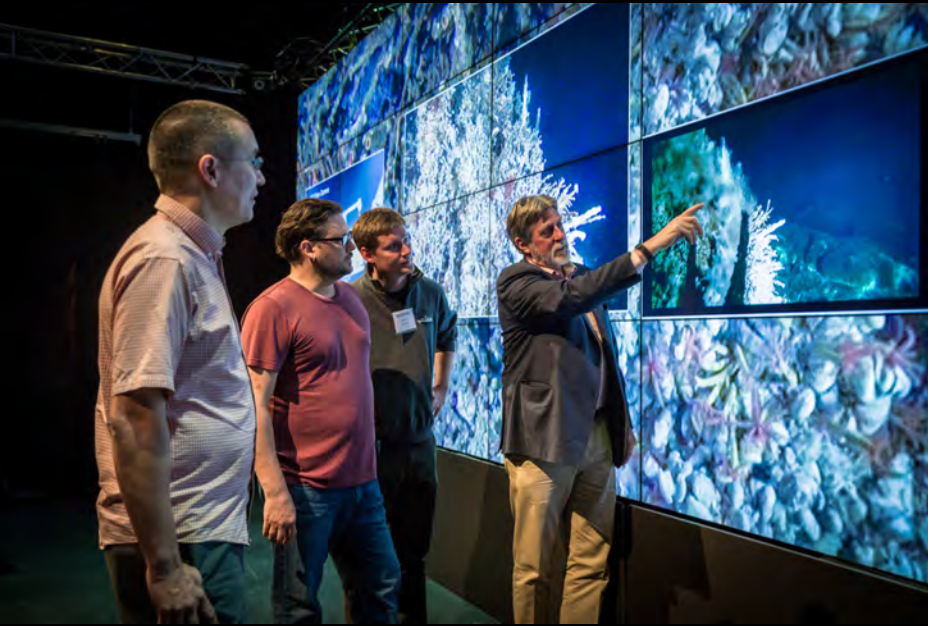
Today, the White House issued a Directive supporting public access to publicly-funded research.

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

appropriate for each type of research conducted or sponsored by the agency."

Beyond Analysis: *Experiencing* Data



Historic Scientific Workflow



Data Acquisition

A computer screen displays a map with blue and green areas. Below the screen is a photograph of a person sitting at a desk with a computer, working. A large blue arrow points from the left towards this box.

Data Processing & Interpretation

A screenshot of a scientific article page from Elsevier. The article title is "Evidence of mass failure in the Hess Deep resolutional bathymetry data". The authors listed are Vicki Lynn Ferrin, Donna J. Shillington, Kathryn G. Damon A.H. Teagle, Antony Morris, and Pierre W. Cazen. The page includes a search bar, navigation options, and social media sharing buttons. A large blue arrow points from the left towards this box.

Publications

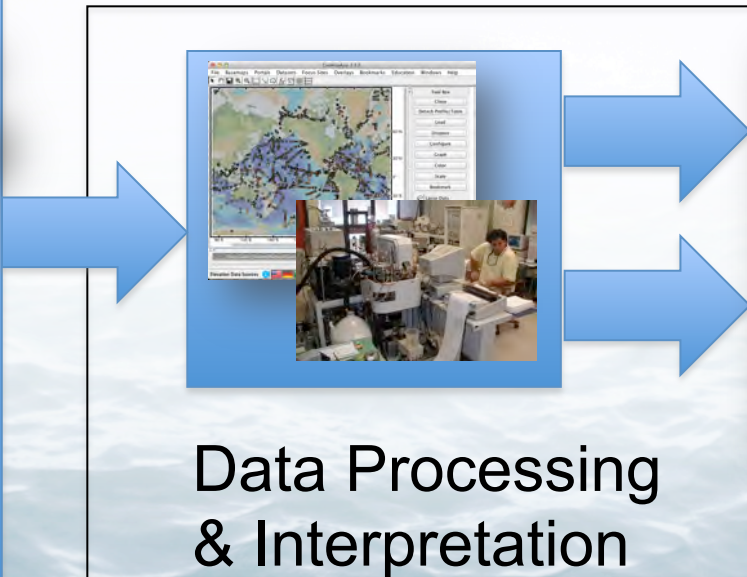
**PRIVATE
PROPERTY
NO PUBLIC
ACCESS**



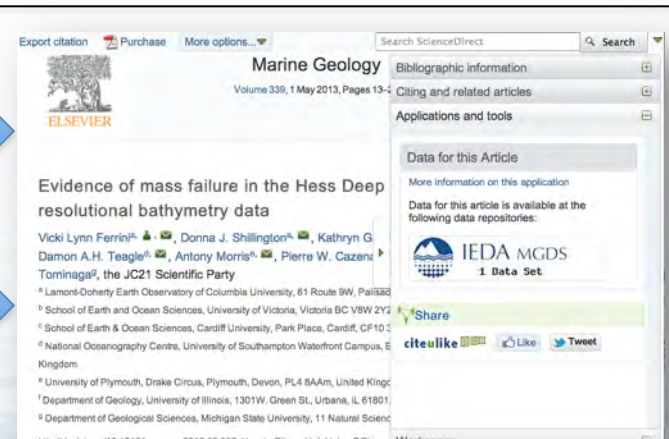
Modern Scientific Workflow



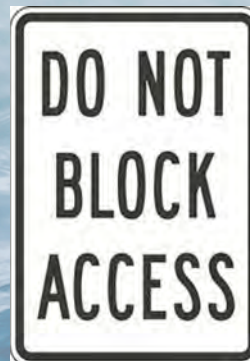
Data Acquisition



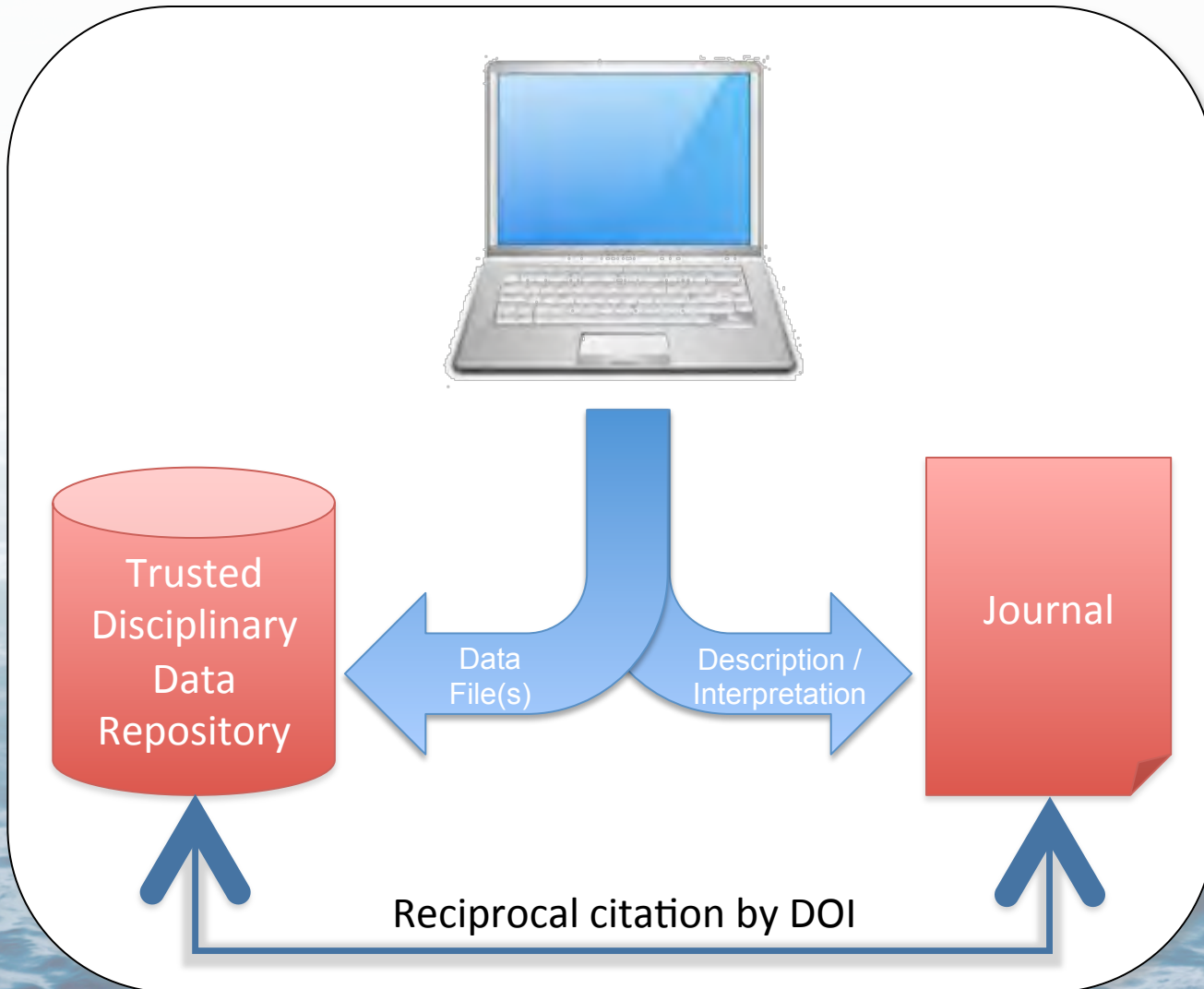
Data Processing
& Interpretation



Publications

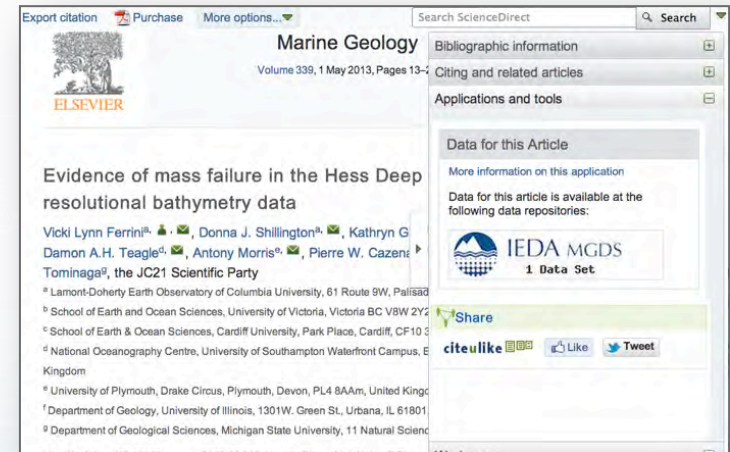


“Best Practice”



What's in it for you?

- Scientific Integrity & Reproducibility
- Collaboration
- Increase Research Impact
- Data Citation
- Compliance with Data Policies
- Outreach & Engagement
- Preserve Data for your own future use

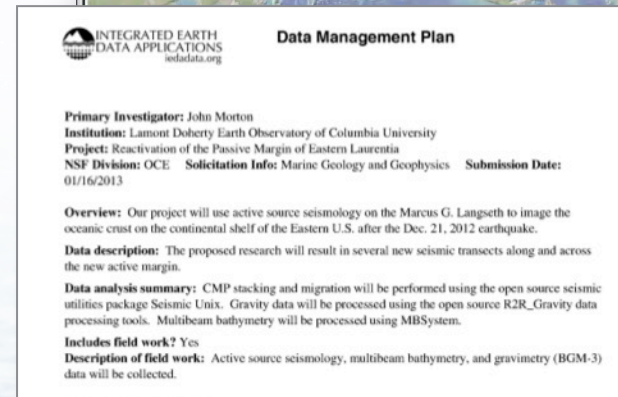
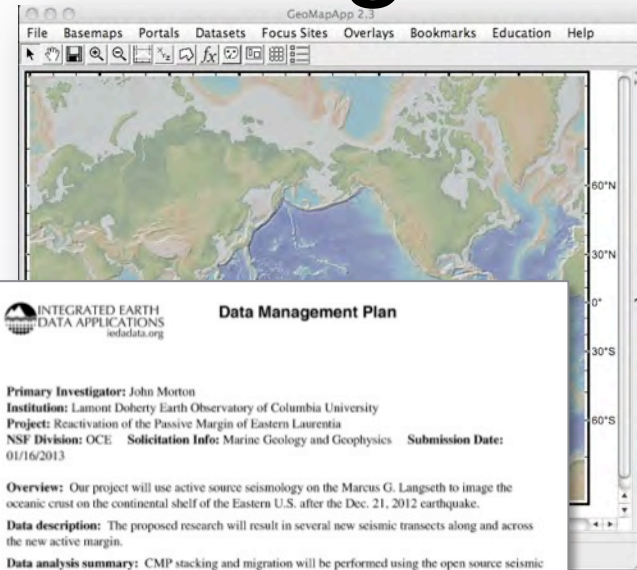


“The coolest thing to do with your data will be thought of by someone else.”

Rufus Pollock

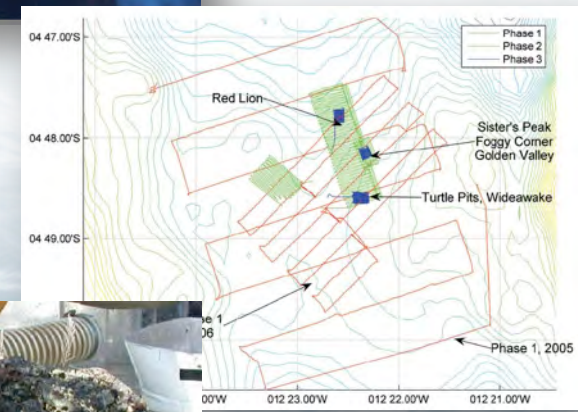
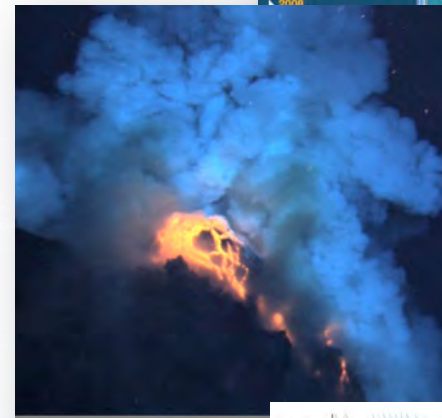
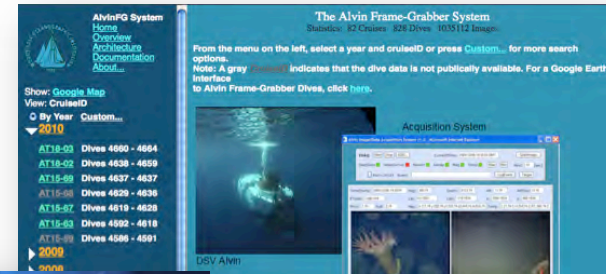
Data Management Planning

- 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?*



Field Data

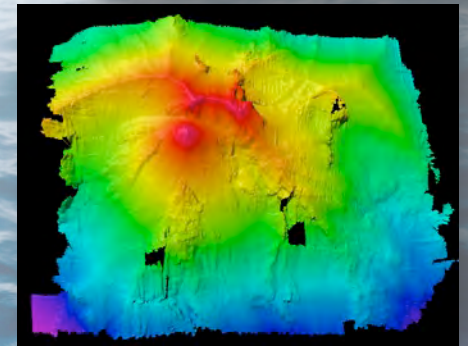
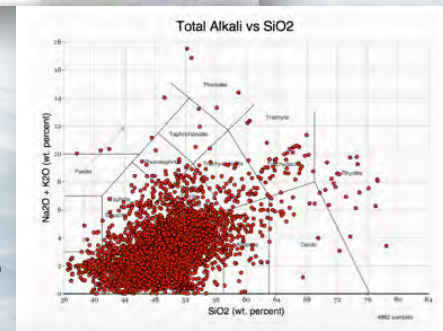
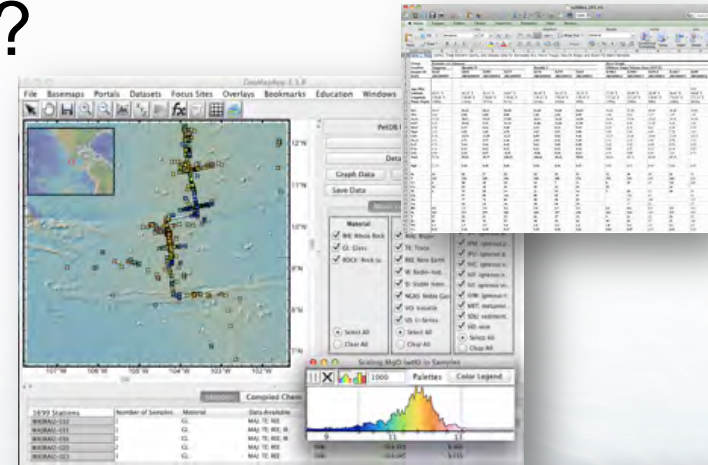
- Facility-Managed Data ✓
 - Ship (R2R)
 - Raw sensor data
 - Vehicle Data (NDSF)
 - Raw sensor data
 - First-order at-sea products
- PI-Managed Data
 - Documentation
 - Cruise report
 - Sample metadata
 - Processing metadata
 - Physical samples
 - Science party instrumentation



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?*





Navigating Data Management

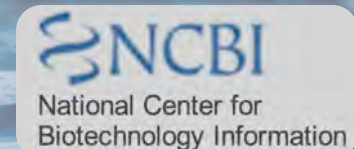
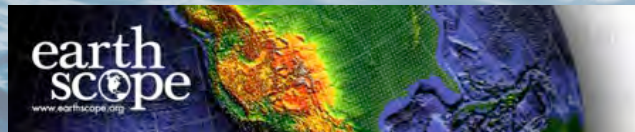
- Plan ahead
- Know what resources are available
 - Software Tools
 - Guidelines & Templates
- Communicate
 - Upstream (Operations Team)
 - Downstream (Data Managers)
- Organize consistently
- Document *contemporaneously*
- Treat data as a valuable community resource
- Participate!
 - Metadata & data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





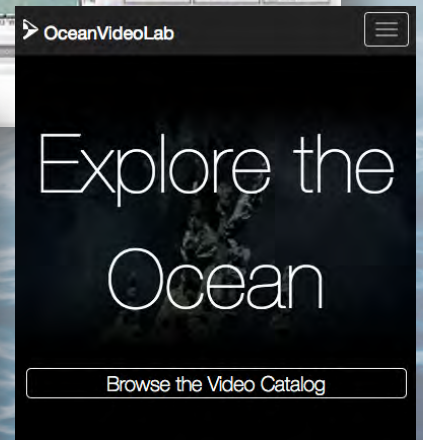
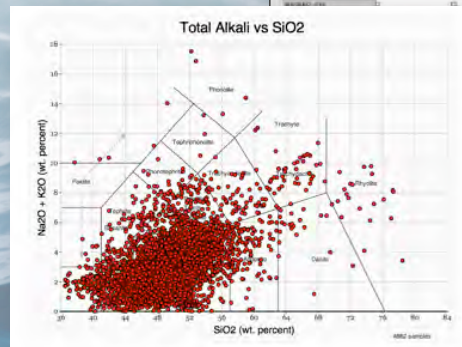
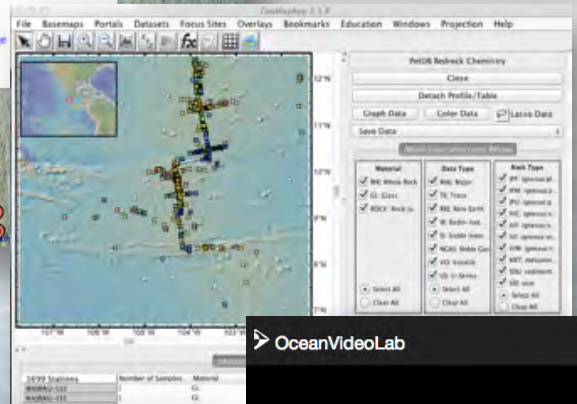
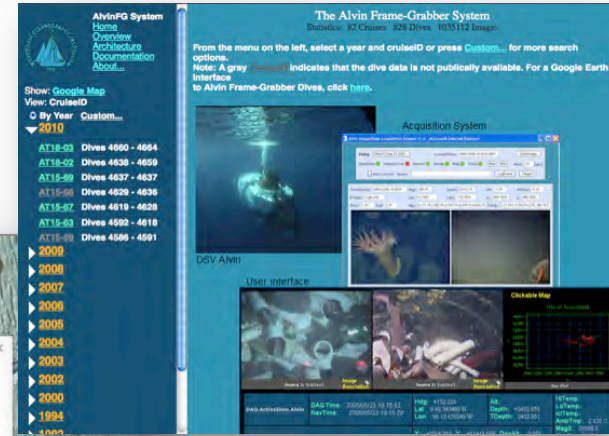
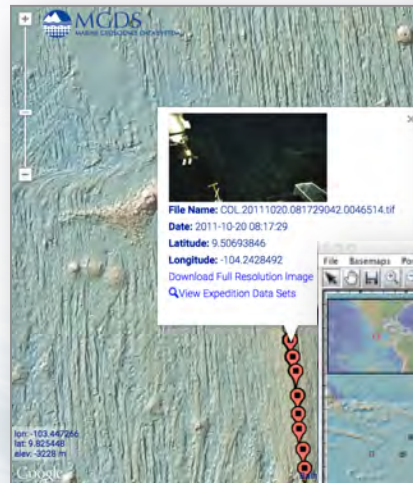
Which Repository?

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



NDSF-Related Data Resources

- 4dgeo.who.edu
 - Jason Virtual Van
 - Alvin FrameGrabber
- iedadata.org
 - EarthChem Library
 - EarthChem Portal
 - GeoMapApp
 - GMRT
 - Marine-Geo Digital Library
 - PetDB
 - SESAR
- oceanvideolab.org



Questions?

ferrini@ldeo.columbia.edu



WOODS HOLE
OCEANOGRAPHIC
INSTITUTION