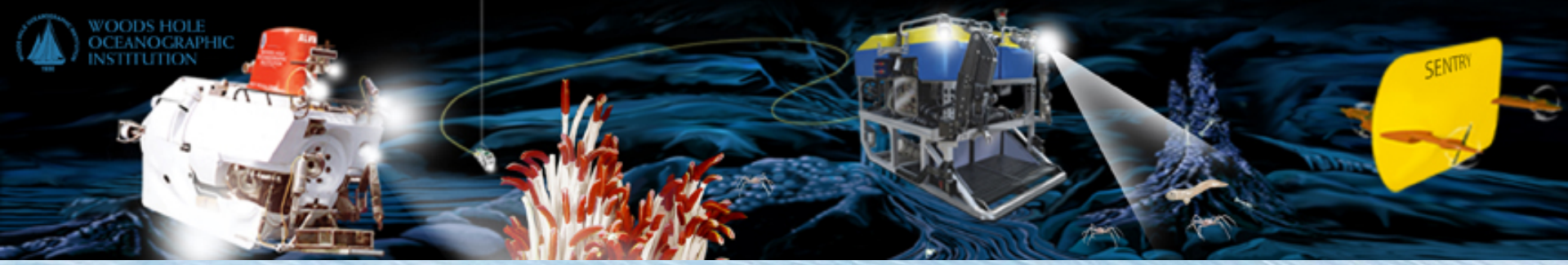


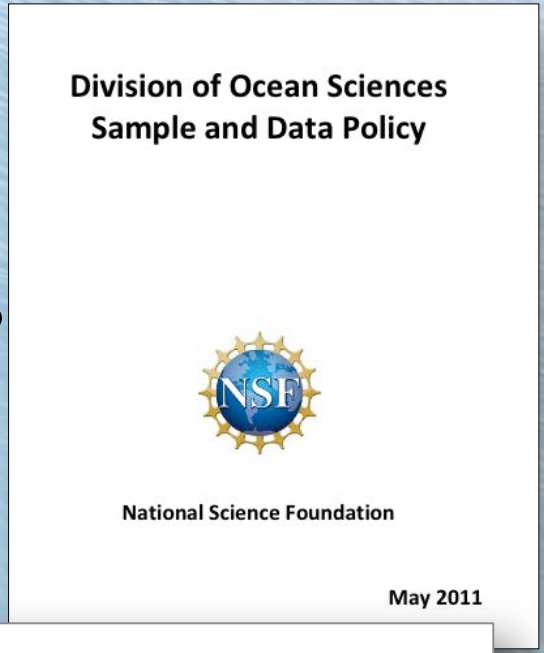
DESSC New User Workshop:
Data Management Overview

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


News

White House issues directive supporting public access to publicly funded research

Timothy Vollmer, February 22nd, 2013

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."

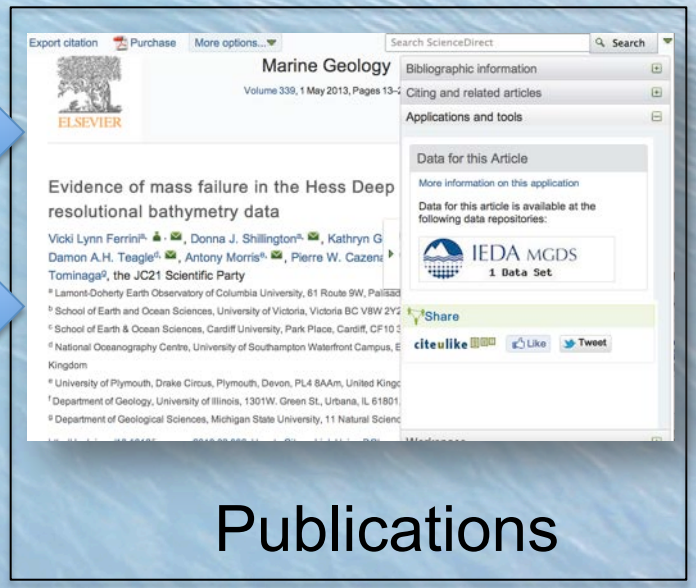
Each agency covered by the Directive (54 KB PDF) must "Ensure that the public can read, download, and analyze in digital form final peer reviewed manuscripts or final published documents within a timeframe that is

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

Historic Scientific Workflow

Data Acquisition

Data Processing
& Interpretation



Publications

PRIVATE
PROPERTY
NO PUBLIC
ACCESS

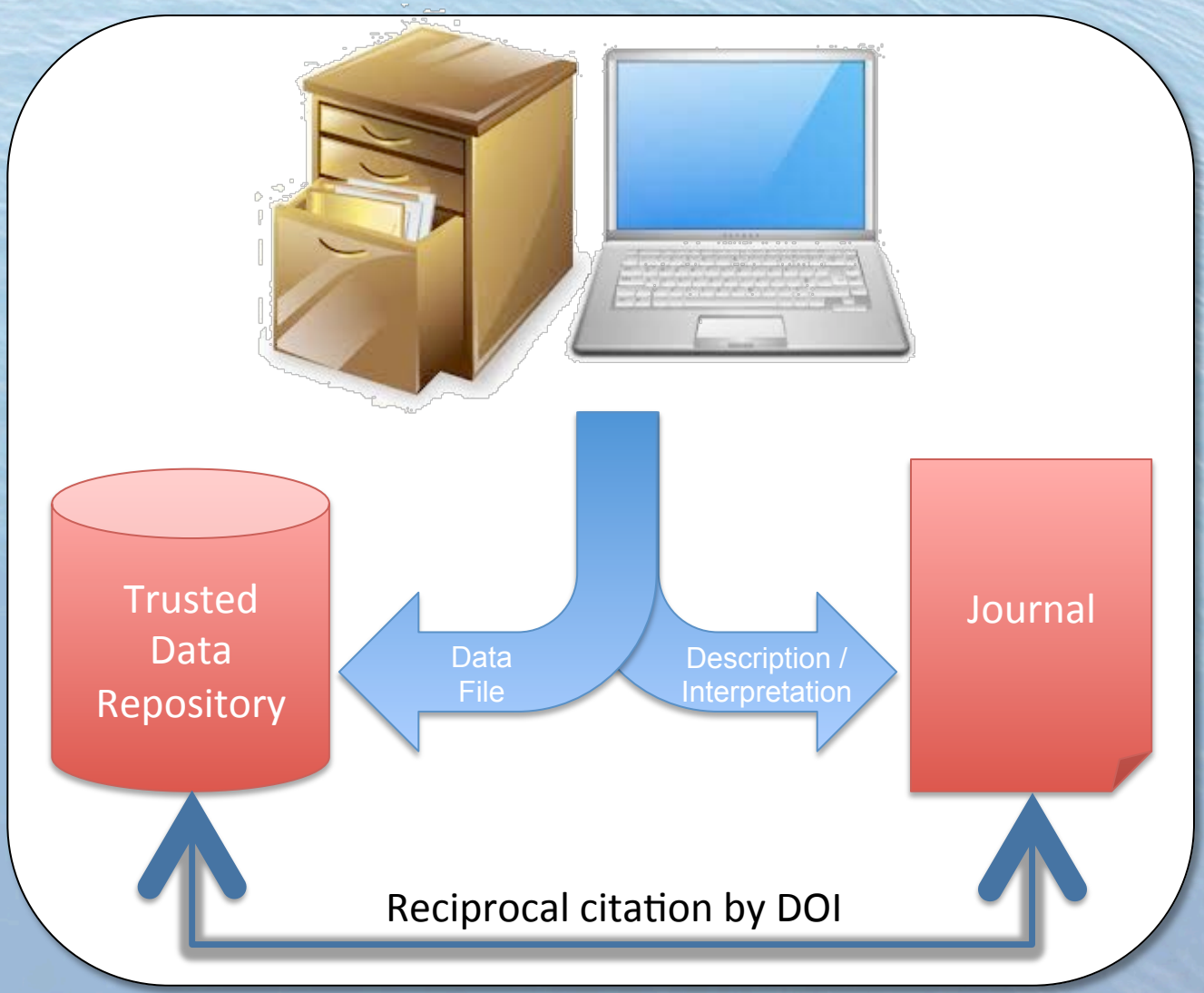
Modern Scientific Workflow

Data Acquisition

Data Processing
& Interpretation

Publications

“Best Practice”



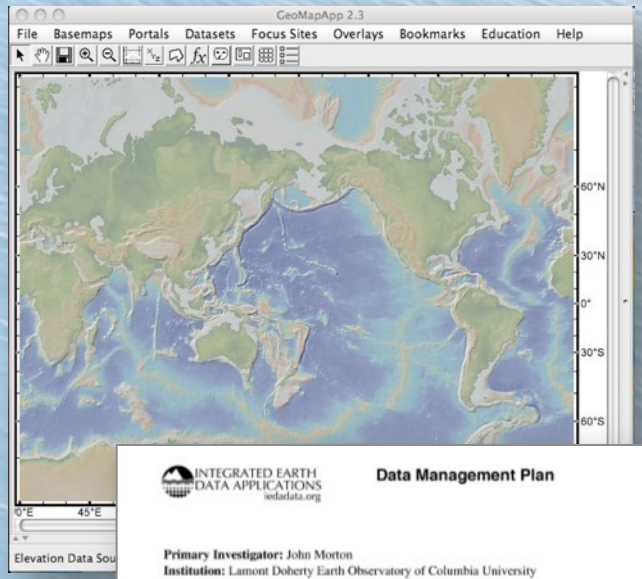
What's in it for you?

- Scientific Integrity & Reproducibility
- 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?*



INTEGRATED EARTH DATA APPLICATIONS
iedadata.org

Data Management Plan

Primary Investigator: John Morton
Institution: Lamont Doherty Earth Observatory of Columbia University
Project: Reactivation of the Passive Margin of Eastern Laurentia
NSF Division: OCE **Solicitation Info:** Marine Geology and Geophysics **Submission Date:** 01/16/2013

Overview: Our project will use active source seismology on the Marcus G. Langseth to image the oceanic crust on the continental shelf of the Eastern U.S. after the Dec. 21, 2012 earthquake.

Data description: The proposed research will result in several new seismic transects along and across the new active margin.

Data analysis summary: CMP stacking and migration will be performed using the open source seismic utilities package Seismic Unix. Gravity data will be processed using the open source R2R_Gravity data processing tools. Multibeam bathymetry will be processed using MBSystem.

Includes field work? Yes
Description of field work: Active source seismology, multibeam bathymetry, and gravimetry (BGM-3) data will be collected.

Expected data product #1
Data type: Observational, Analytical
Responsible investigator: John Morton
Product description: .segy files from seismic transits.
Intended repository: IRIS
Timeline for data release: Immediate Release

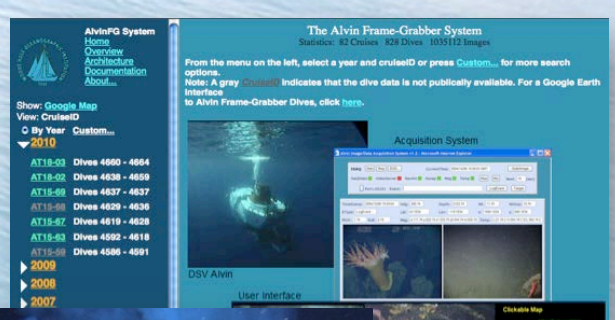
Expected data product #2
Data type: Observational
Responsible investigator: John Morton
Product description: Processed free-air anomaly data in MGD77-T format
Intended repository: NGDC
Timeline for data release: Immediate Release

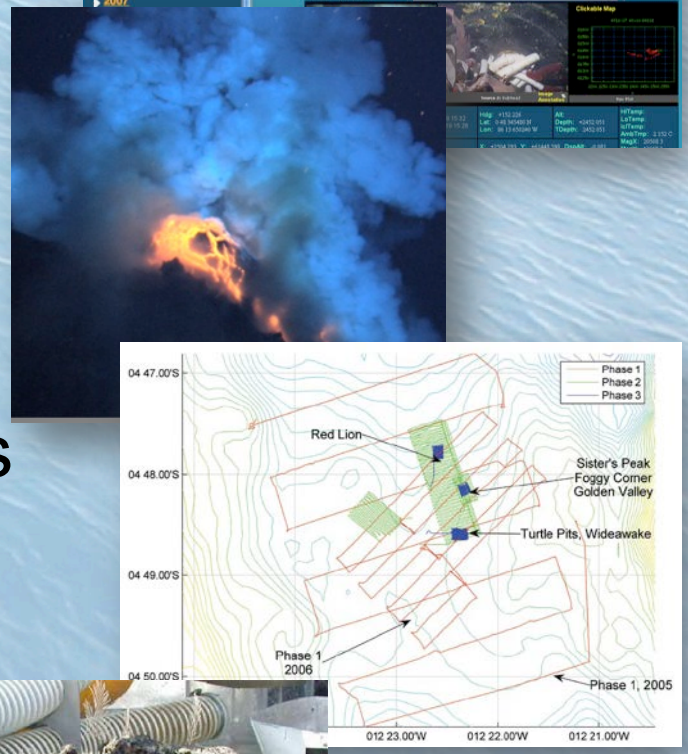
Expected data product #3
Data type: Observational
Responsible investigator: Vicki L. Ferrini
Product description: Multibeam bathymetry data
Intended repository: MGDS
Timeline for data release: Immediate Release

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

Field Data

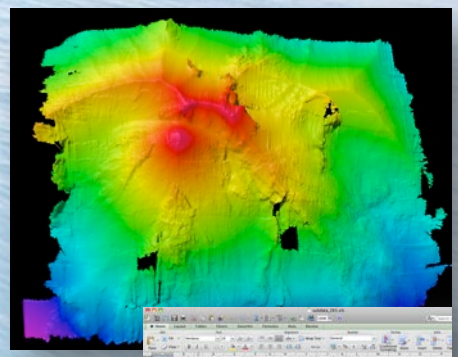
- Facility-Managed Data ✔
  - Ship (R2R)
    - Raw sensor data
  - Vehicle Data (NDSF)
    - Raw sensor data
    - First-order at-sea products
- PI-Managed Data
  - Documentation
  - Cruise report
  - Sample metadata
  - Processing metadata
  - Physical samples
  - Science party instrumentation

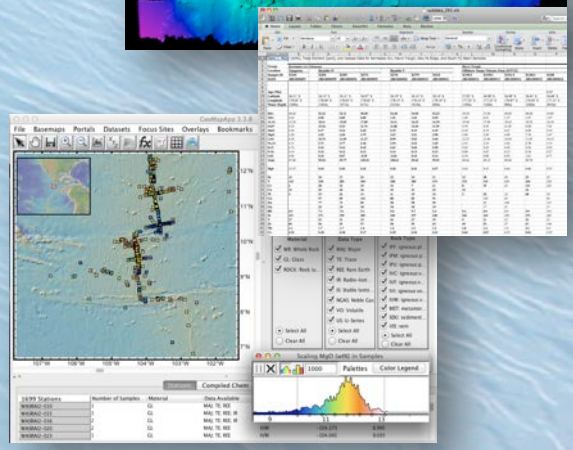


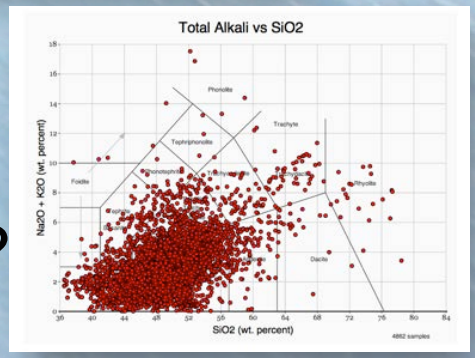


Derived Data

- 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?*





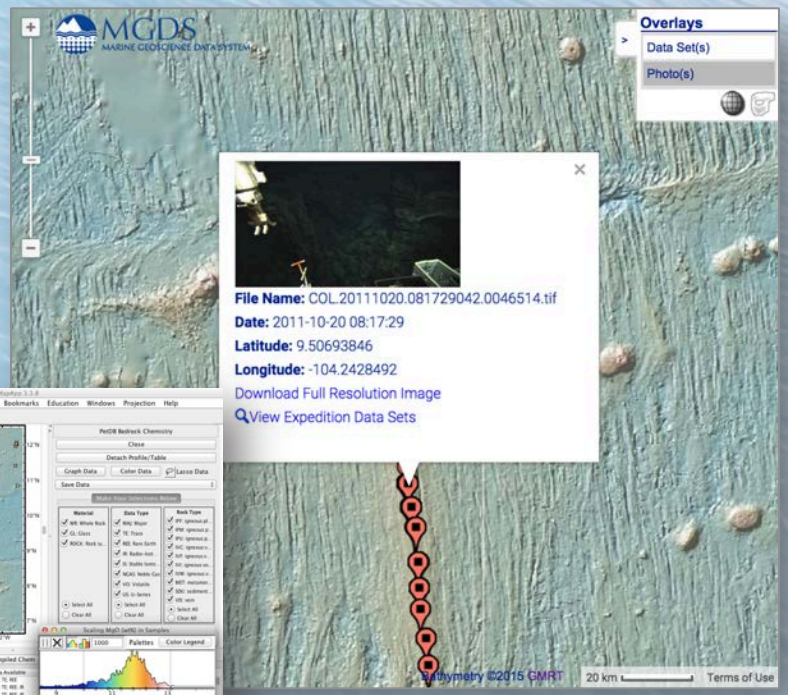


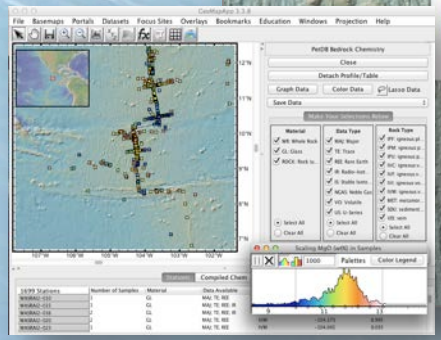
Navigating Data Management

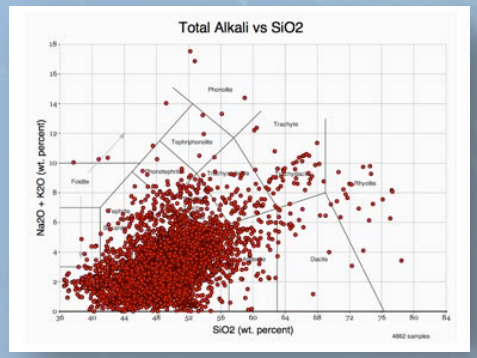
- Plan ahead
- Know what resources are available
  - Tools to make process easier
- Communicate
  - Upstream (Operations Team)
  - Downstream (Data Managers)
- Document *contemporaneously*
- Treat data as a valuable community resource
- Participate! Science community input needed for:
  - Metadata & data format standards
  - System usability

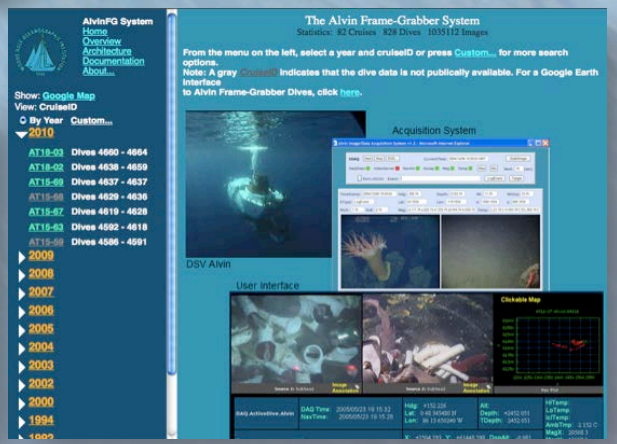
NDSF-Related Data Resources

- http://4dgeo.whoi.edu
  - Jason Virtual Van
  - Alvin FrameGrabber
- http://www.iedadata.org
  - EarthChem Library
  - EarthChem Portal
  - GeoMapApp
  - GMRT
  - Marine-Geo Digital Library
  - PetDB
  - SESAR









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

IGSN e.v.

IEDA
INTEGRATED EARTH DATA APPLICATIONS

rvdata.us

MGDS
MARINE GEOSCIENCE DATA SYSTEM

PetDB Petrological Database

Explore our planet with
GeoMapApp

IRIS

NCAR

UNAVCO

unidata

BCO-DMO

NCBI
National Center for
Biotechnology Information

IODP
INTEGRATED OCEAN
DRILLING PROGRAM

CSDMS
COMMUNITY SURFACE DYNAMICS MODELING SYSTEM

National CZO Program
CRITICAL ZONE OBSERVATORIES
funded by NSF
studying the zone where rock meets life

earthscope

NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE

NASA

USGS
science for a changing world

Questions?

ferrini@ldeo.columbia.edu