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

Enables New Opportunities:
- Spatial & temporal change
- Scientific reproducibility
- Data syntheses
- New possibilities for analysis
Beyond Analysis: Experiencing Data
Historic Scientific Workflow

Data Acquisition

Data Processing & Interpretation

Publications

PRIVATE PROPERTY
NO PUBLIC ACCESS
Modern Scientific Workflow

Data Acquisition

Data Processing & Interpretation

Publications

DO NOT BLOCK ACCESS
“Best Practice”

Trusted Disciplinary Data Repository

Data File(s)

Description / Interpretation

Journal

Reciprocal citation by DOI
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
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?
Field Data

- Facility-Managed Data ✓
  - Ship (R2R)
    - Raw sensor data
  - Vehicle Data
    - Raw sensor data
    - First-order at-sea products
- PI-Managed Data
  - Documentation
  - Cruise report
  - Sample metadata
  - Processing metadata
  - Physical samples
  - Science party instrumentation
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 ([whoi.edu](http://whoi.edu))
  - Dive metadata & data at WHOI Data Library
  - NDSF vehicles
- Marine Geoscience Data System ([marine-geo.org](http://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, ABE, Sentry
    - MBARI Mapping AUV D. Allan B.
    - REMUS
    - Legacy: DSL120, IMI130
- NOAA/NCEI ([www.ncei.noaa.gov](http://www.ncei.noaa.gov))
  - OER dive metadata & data
  - Searchable dive video archive
- Ocean Video Lab ([oceanvideolab.org](http://oceanvideolab.org))
  - Video annotation leveraging YouTube
  - Modern and legacy video
- 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
• 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

Lightning Talks: AGU How to Find an Appropriate Repository
Thursday 12/12, 11-11:45 AM
Exhibit Hall Booth # 1329
Questions?

Come find me at the DeSSC Meeting tomorrow!

-or-

Email me any time: ferrini@ldeo.columbia.edu