IEDA
iedadata.org

Supporting Data Stewardship Throughout the Data Life Cycle in the Solid Earth Sciences

Vicki L. Ferrini, Kerstin A. Lehnert, Suzanne M. Carbotte, and Leslie Hsu

Lamont-Doherty Earth Observatory

IEDA
iedadata.org

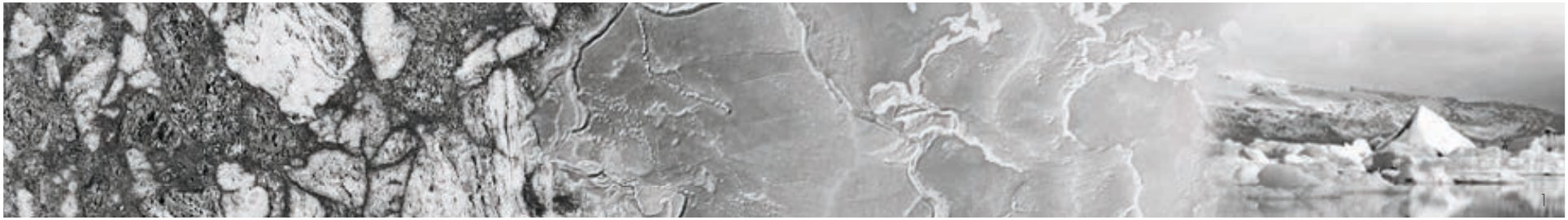
What is IEDA??

A community-based data facility funded by NSF to support, sustain, and advance the geosciences by providing data services for observational solid earth data from the Ocean, Earth, and Polar Sciences.

- Domain-specific Data Libraries
- Data Syntheses
- Visualization & Analysis Tools
- Data & Metadata Registries
- Data Publication
- Investigator Support Services

MARINE GEOSCIENCE DATA SYSTEM

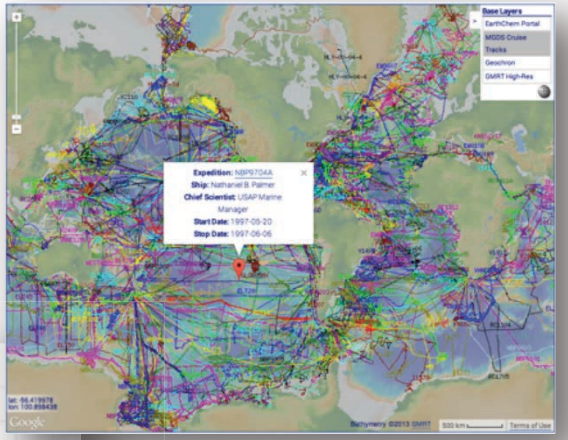
EarthChem



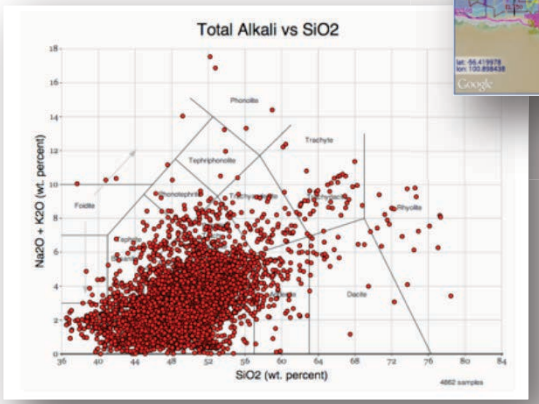
IEDA
iedadata.org

Data Curated in IEDA Systems

- Sensor-based Data
  - Marine geophysical data (e.g. bathymetry, sidescan, subbottom)
  - Complementary datasets
    - Navigation, bottom photos
    - Technical reports



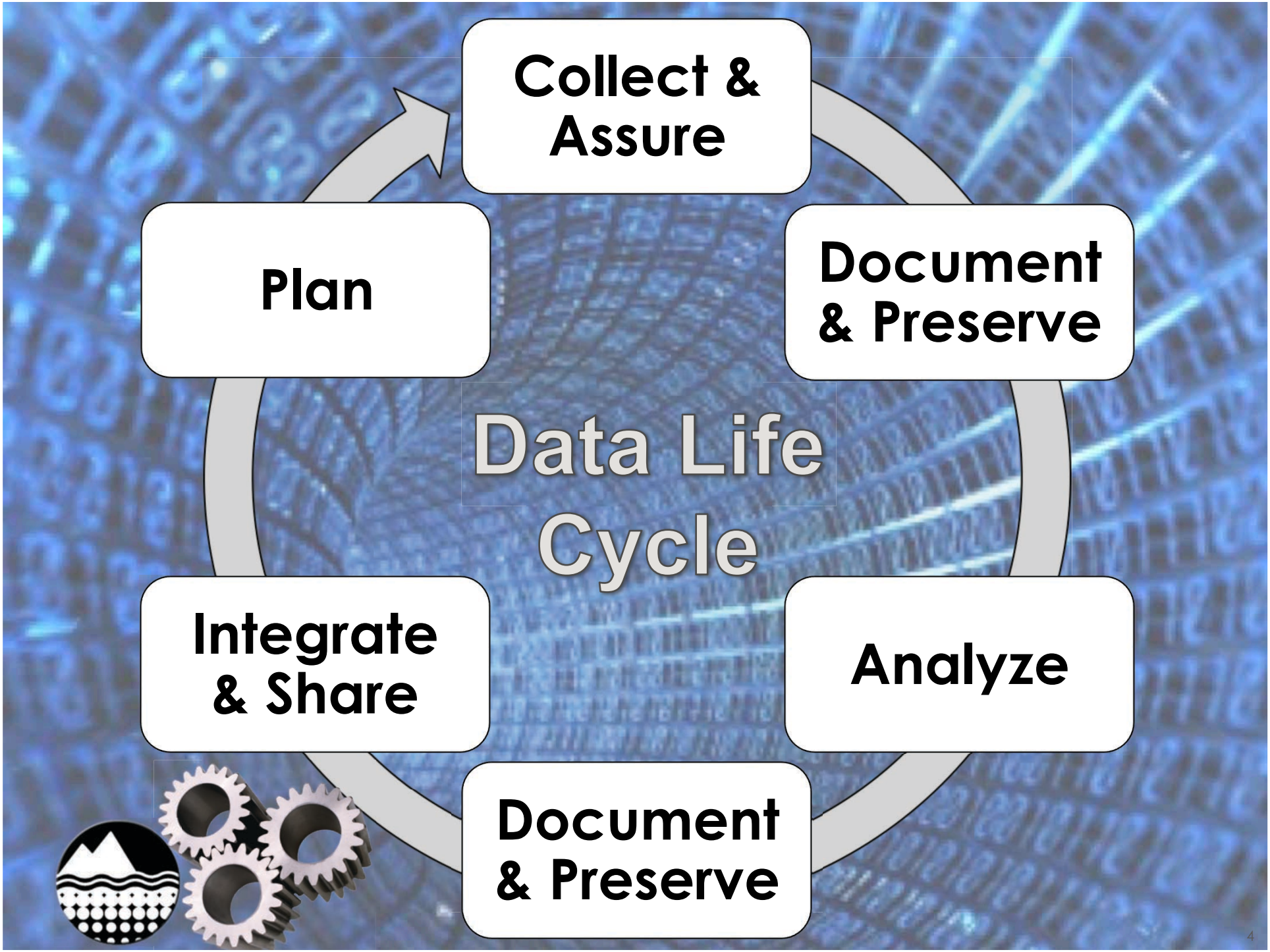
- Sample-based Data
  - Sample metadata
  - Analytical results
  - Technical reports



IEDA
iedadata.org

Domain-specific Data Curation

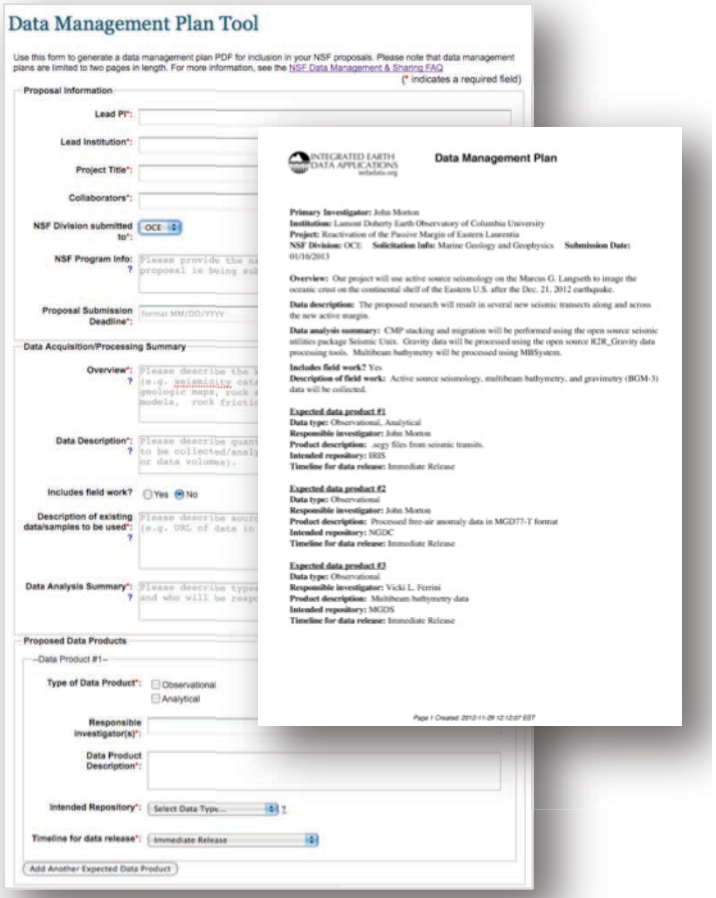
- Familiarity with content, acquisition, processing, & use of data
- Ability to evolve services in response to both:
  - User-community needs
  - Technology
- Construct & curate high-quality syntheses
- Curate rich metadata
  - Enhanced discoverability & usability
  - "Fitness for re-use"


Collect & Assure
Document & Preserve
Analyze
Document & Preserve
Integrate & Share
Plan
Data Life Cycle

IEDA
iedadata.org

Data Life Cycle: Plan

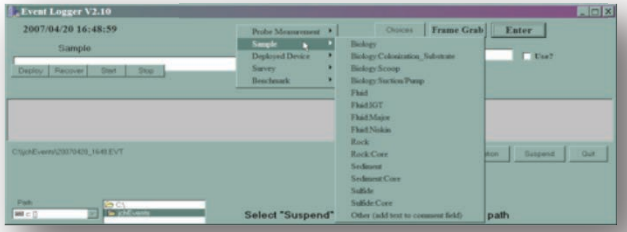
- Data Management Plan Tool
  - Facilitate assembly
  - Inform Investigators
  - Inform down-stream repositories
  - Promote dialogue
- Data Acquisition Plan
  - Metadata & data templates
  - Promote & facilitate *contemporaneous* documentation

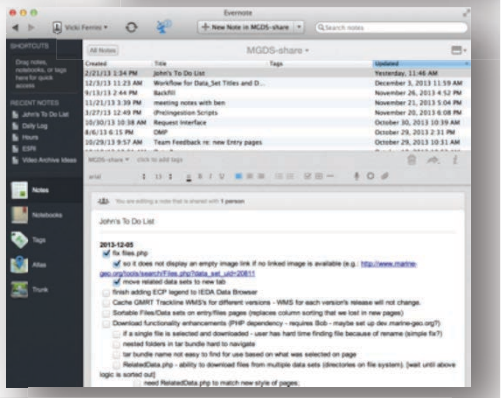


IEDA
iedadata.org

Data Life Cycle: Collect & Assure

- Promote Best Practices
  - What to document
  - How to document
- Tools and workflows to facilitate digital documentation
  - Metadata & Data Templates



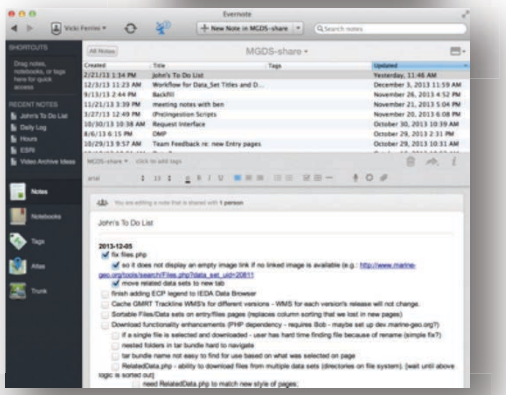


IEDA
iedadata.org

Data Life Cycle: Document & Preserve

- Document & capture data & metadata as soon as it is available
  - Simple interfaces & guidelines
- Sample metadata registry
- Link to complementary data & metadata

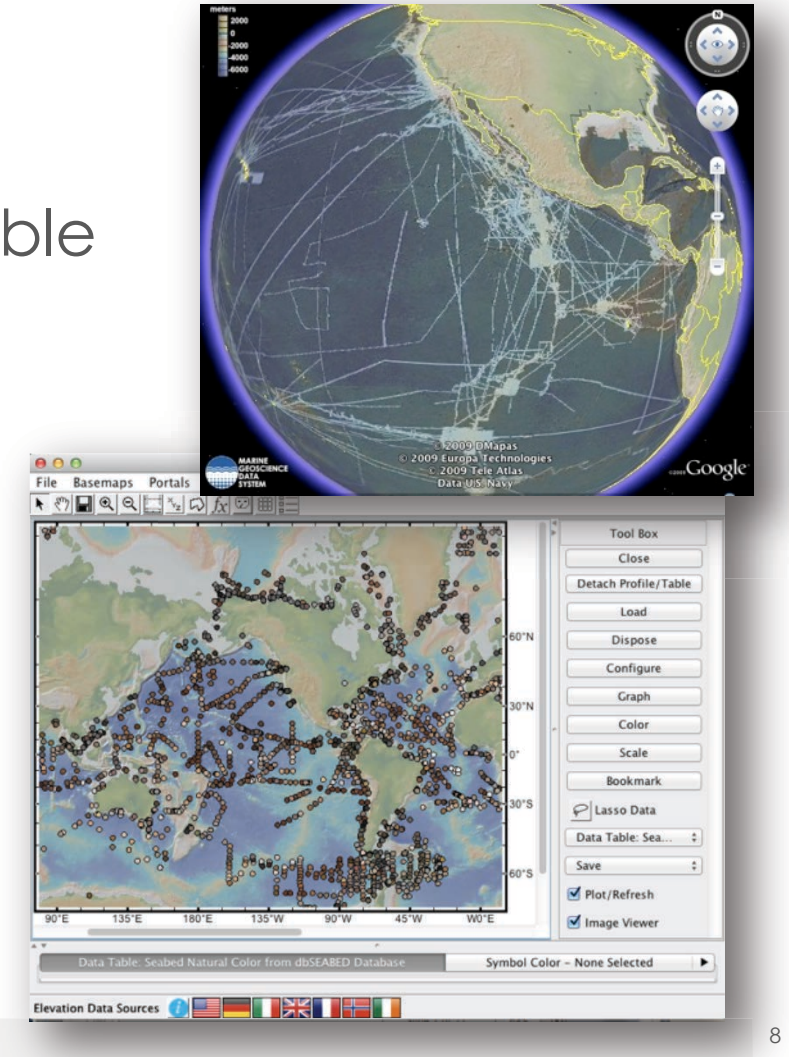
IGSN e.v.



IEDA
iedadata.org

Data Life Cycle: Analyze

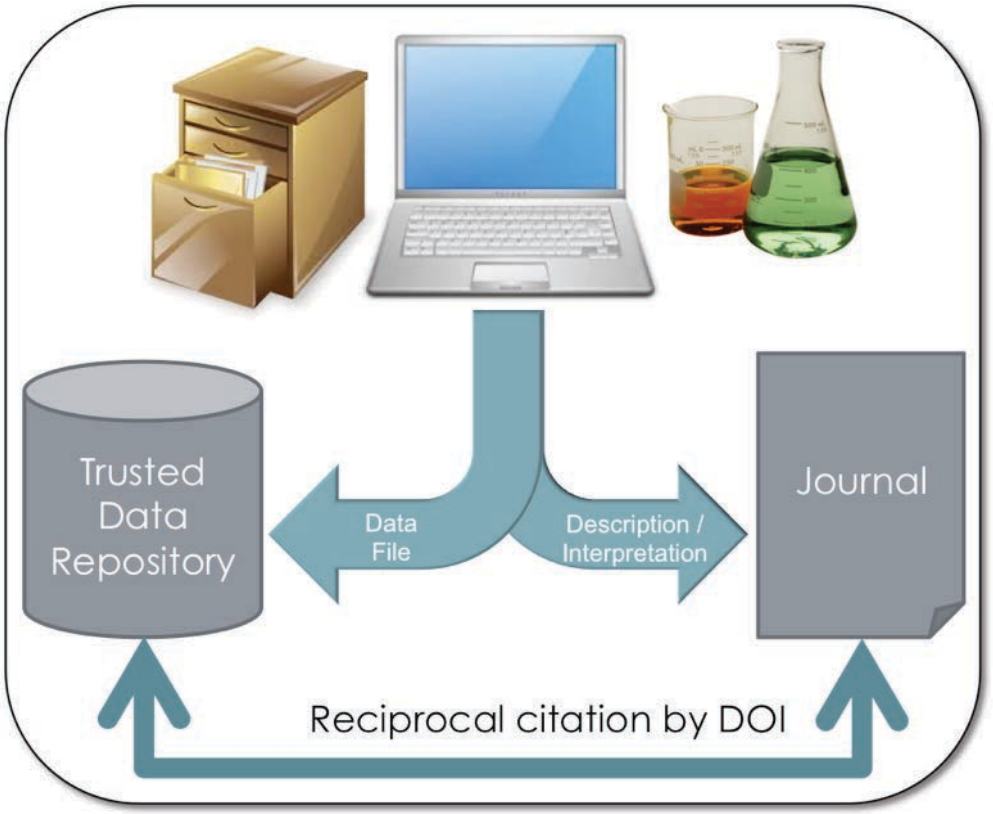
- Tools to:
  - Support domain specialists
  - Make specialist data accessible to non-specialist users
  - Integrate & visualize
- Access to Data Syntheses
- Access to complementary data & resources



IEDA
iedadata.org

Data Life Cycle: Integrate & Share

- Advise on what to preserve & how
  - Data supporting pubs
  - Data of value
- Facilitate data prep.
  - Metadata requirements
  - Templates
  - Format guidelines

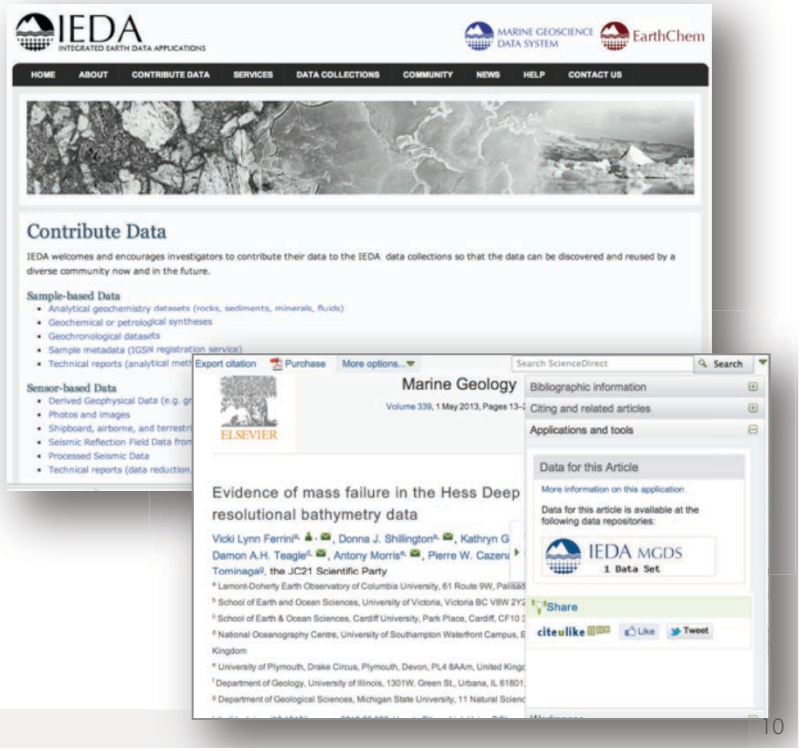


IEDA
iedadata.org

Data Life Cycle: Document & Preserve

Develop simple workflows, interfaces & templates to capture sufficient information for:

- Long-term curation & access
  - Inclusion in syntheses
- Links to scientific publications
- Data Publication
- Data use, discovery & re-use
- Attribution & collaboration
- Data Download Stats
- Data Compliance Reporting

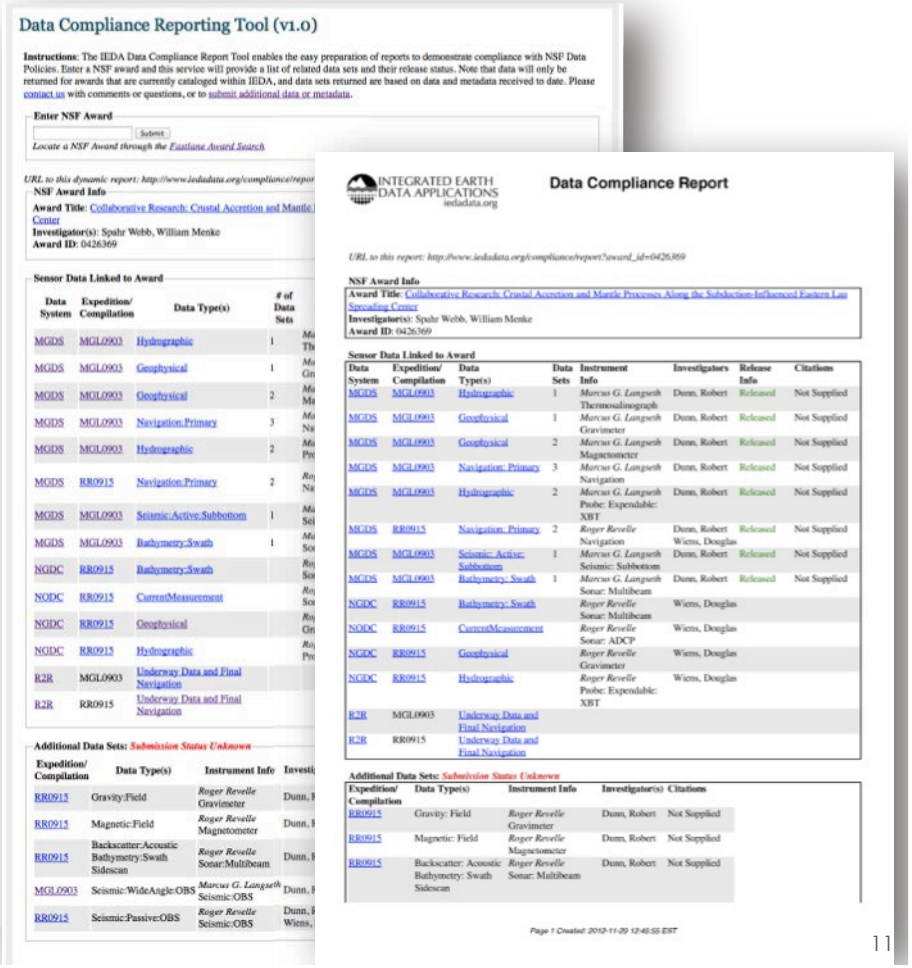


IEDA
iedadata.org

Data Compliance Reporting Tool

- Tool for demonstrating compliance:
  - Award-based
  - Informed by DMP
- Report includes:
  - Data Inventory
  - Release Status
  - Links to data
  - Save as PDF

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



IEDA
iedadata.org

Summary

- Domain-specific repositories have an important role to play throughout the data life cycle:
  - Increase awareness of resources & tools
  - Lessen "burden" of data management
  - Provide guidelines & tools for data curation, discovery & access
  - Respond to evolving needs of user community
- Greatest success when repository team members are integrated members of user community

IEDA
iedadata.org

Opportunities & Challenges

- Data management training
  - Minor modifications throughout life-cycle will yield great results
  - Balance between data management & science training
  - Barriers to adoption must be low
- Evolving technologies can be leveraged
  - Pace of technology vs. pace of adoption
- Evolving significance of data publication
  - Minimal metadata approaches are easy but less useful