Integrated Data Management System for Marine Geoscience Data

Scope of the Integrated DMS
- Sensor Database: multibeam and geophysical data from Palmer & Gould and MCS data from Ewing & Langseth
- Metadata catalog: Central cruise catalog and data repository for these ships as well as for all R2K and MARGINS programs
- Global DEM: Synthesis of global bathymetry into multi-resolution DEM
- MG&G Legacy data
- Tools for data access: lower barrier to data access with tools tailored to science needs

Design Requirements
- Generalized Data Model: Need to handle large and diverse datasets (seismic, sonar, geological, fluid, biological, rock, and sediment samples, temperature, photo imagery)
- Easy to Use: Need to serve a diverse and multi-disciplinary user community - non-specialist and specialist access
- Enable New Discovery: Global coverage to facilitate global syntheses
- Optimize Collective Resources: Link to and integrate with existing data resources
- Community Guidance: Remain closely tied to science user community throughout design and operations to ensure optimal utility

MGDS Funding History

<table>
<thead>
<tr>
<th>Year</th>
<th>Legacy (ITR)</th>
<th>Seismic Reflection DMS</th>
<th>MARGINS DMS</th>
<th>Ridge2000 DMS</th>
<th>Antarctic MBS (OPP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Work Tasks
- Shared Infrastructure:
  - Database design, development, and review
  - Application design, development, implementation
  - Web site development and maintenance
  - Interoperability development - Web services, controlled vocabularies
  - Data solicitation and QA procedures
  - Global digital elevation model
  - System maintenance [hardware, backup, upgrades, security, administration]
Work Tasks (cont.)

Project Specific:
- Solicitation of data
- Metadata and data entry
- Legacy Data Recovery
- Outreach and community liaison
- Education

Organizational Structure

Project Management

Web Accessible Content Management System - Plone

User Support

Request Tracker

Educational Activities

- Student research projects ([5 high school, undergraduate and graduate students involved in summer 2006])
- Earth2Class Teacher Training
- User training - Hamilton College
- MARGINS EU Wksp
- R2K EU Component
Metrics of Success

- Tracking Use

Guidance and Outreach

- Biannual updates at R2K and MARGINS STCOMM
- Reports in R2K/MARGINS newsletters/Pamphlets
- Annual Advisory Committee Meeting
- Exhibits at national meetings (AGU)
- Numerous community workshops

Relationship with broader data infrastructure for geosciences

- Community Collection
- Reference Collections
- Data Documentation
- Data Validation
- Data Integration
- Data Discovery
- Data Access
- Data Synthesis

Relationship to Data Collectors

- Raw data
- Metadata
- Metadata forms
- QA feedback

Relationship to Data Users

- Global DEM
- Data for Users
- Map-based tools
- Derived data, feedback
- Planning/priorities
Metadata, field data, QA MB data

MGDS-NGDC

MGDS

Author FGDC records

Metadata extensions, controlled vocab

Data provider

Advisory Structure: Current Model

Advisory Structure: Model 2

Advisory Structure: Model 3

Joint GIG-MGDS Advisory Committee

Sub-Committee
e.g. R2K/MARGINS STCOM

Sub-Committee
e.g. EarthChem AC
Managing Digital Data Collections for the Earth & Ocean Sciences

Project Execution & Advisory Structure
Digital Data Collections

“resemble large facilities projects” [NSB Report]

in terms of

- their extended lifetime,
- the need for stable core support,
- the critical importance of effective project management in combination with domain expertise,
- the ability to energize and enable broad research and education communities, and
- the importance of partnerships, both national and international.”
Responsibilities

- Reliable & competent partner in data archiving & preservation
- Participate in the development of community standards (metadata, data format, DQC)
- Ensure that standards are universally applied to data
- Provide for integrity, reliability, and preservation of the collection by developing & implementing plans for backup, migration, maintenance, and all aspects of change control
- Provide for security of collection
- Provide contextual information including cross-references to other data sources
Data Systems as Community Resources

- Accountability toward community
- Accountability toward funding agency

“Although the LDEO activities are separate projects, the number of these has grown to the extent that a separate management effort is probably now required, not only at LDEO, but at NSF as well.”

*NSF, Oct 2005*
Project Execution Plan

- Describes strategic and management approaches, organizational structures and timelines under which the projects will be executed.

- Is intended as a dynamic document, reviewed and revised annually in discussion with the Advisory Committee and the National Science Foundation.
  - Revisions will reflect changes in the scope and type of projects included, in the organizational structure, and in the funding.
  - New projects will be integrated into the plan as appropriate.
PEP Components

- Project Definition
- Organizational Structure
- Project Management
  - Roles & Responsibilities
  - Work Breakdown Structure
  - System Development & Operation Strategies
  - Risk & Contingency Management
  - Timelines
- Project Monitoring & Assessment
  - Reporting
  - Metrics of Success
  - Advisory Structure
Focus: Synergies

- Exist both within and among the two project clusters
  - data types, science applications, user communities, and technical requirements
- Offer extensive opportunities for collaboration & integration
- Result in substantial ‘Economies of Scale’
- Positively impact the quality of the products
- Rationalize the development and operation of various systems by an integrated experienced team
Synergies

SYNERGIES
Lamont Digital Data Collections for Geochemistry and MG&G

Metadata Standards
Education Activities
Science Applications
Visualization Tools (GeoMapApp)

Geoinformatics for Geochemistry Program
- System engineering
- Systems Op and Maintenance
- Data modeling (samples)
- Implementation of IGSN
- Map Interfaces
- Community Outreach

Geochemistry Data Collections
- Data modeling (geochemistry)
- Data submission & QC procedures
- Interoperability (XML, web services)
- Query interfaces (SQL, Java)

Marine Geoscience Data System
- System engineering
- Systems Op and Maintenance
- Data Submission procedures
- Interoperability (Web Services, metadata)
- Query interfaces
- Data Visualization Tools

PetDB, SedDB, Earth Chem, SESAR

Project-specific tasks
- Data solicitation and QC
- Data entry & loading
- Custom Query interfaces
- User support
- Community Outreach
Examples

- application of GeoMapApp as a map interface and plotting tool for the geochemical data systems,
- use of sample metadata schemes developed by PetDB for the MGDS metadata catalog,
- seamless data exchange between PetDB and the MGDS,
- use of MGDS cruise metadata for SESAR sample profiles,
- Use of MGDS bathymetry for EarthChem maps
PEP Implementation

▶ Create the Geoinformatics for Geochemistry Program as a new integrated management structure for the geochemistry data collections.

▶ Seek required funds

▶ to support the implementation of the GfG program.
▶ to extend joint education activities by the MGDS and the GfG program.

Proposal submitted August 2006, pending

Funds to support the development of the MGDS as a fully integrated system are fully supported by the existing grants.
Proposal Components

- Implementation of the GfG Program
  - Salary support for program coordination and project management
  - Data stewardship

- Joint educational activities
  - Summer Institute for Geoinformatics
    - configured in a way that it improves mutual understanding among geoscientists and technology experts.
    - series of Geoscience, Geoinformatics, and Information Technology lectures by LDEO and CIESIN staff and, if possible, by participants of the summer institute
    - joint exercises that foster interaction, exchange, and understanding between participants of technical and Geoscience background
    - one or more field trips to educate participants with hands-on experiences about field and lab methodologies
  - Summer Internships
    - Part of LDEO Summer Intern Program (D. Abbott)
Future Direction

△ Primary goal is to ensure the highest level of quality and long-term sustainability of systems, data products, & the services they provide to the community.

△ Maintain a level of funding to support operations with the existing superior expert teams.
△ Continue infusion of relevant new technologies.
△ Focus on expanding collaborations with other Geoinformatics efforts.
△ Contribute to establishing a cyberinfrastructure for Geoscience research and education.
The GfG Program

► a new integrated organizational structure for the data management projects carried out by the LDEO-CIESIN collaboration
GfG Program: Rationale

- independent projects share lots of synergies
- carried out by the same group of people
- Increased Level of Effort demands increased level of project management
  - Project coordination
  - Project monitoring and assessment
  - Accountability & validation
  - Community involvement & education
- Initiate system integration
Funding periods

<table>
<thead>
<tr>
<th></th>
<th>FY03</th>
<th>FY04</th>
<th>FY05</th>
<th>FY06</th>
<th>FY07</th>
<th>FY08</th>
<th>FY09</th>
<th>FY10</th>
<th>FY11</th>
</tr>
</thead>
<tbody>
<tr>
<td>PetDB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SedDB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EarthChem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SESAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Synergies among GfG Projects

- Combined experience of team
- Requirements and specifications development
- Database design
- Data preparation, validation, and ingest
- Applications and tools development
- Interoperability
- Outreach & education activities
Examples for Synergies

▷ Generic data model GCDM
  ▶ originated in database design activities for SedDB
  ▶ will be used for EarthChem and potential future databases (VentDB)

▷ Data loading spreadsheets and procedures
  ▶ developed for PetDB
  ▶ adopted for SedDB and EarthChem
  ▶ significantly helped the development of flexible and modular data validation and ingest tools that can be adapted to various projects with some customization

▷ XML schema and web services
  ▶ developed for publishing geochemical data as part of the EarthChem project
  ▶ will promote interoperability and data sharing in a standard manner across all projects

▷ OGC compliant mapping services
  ▶ being prototyped for the PetDB project
  ▶ can be implemented for the other projects with little customization
  ▶ will promote serving sample metadata and data via interoperable interfaces
GfG: A PROGRAM

- “collection of organizational resources that is geared to accomplish a certain major goal or set of goals”
- layer above project management aims to provide an infrastructure where the individual projects can be run successfully
- Program management focuses on coordination among the projects in order
  - to maintain and implement the common program vision,
  - to identify and take advantage of synergies and economies of scale across the sub-projects,
  - to resolve scheduling and resource conflicts.
GfG Program: Vision

- provide for the long-term curation of scientifically important geochemical and related data;
- develop a broad knowledgebase that supports discovery and use of data holdings across projects;
- support consistent best-practices and leading edge scientific data management across all projects;
- leverage the expertise, experiences, and ongoing training of project staff to educate the developing cadre of professional data managers as well as domain scientists and graduate students in GI methodologies and approaches
GfG Program: Vision

- provide for a rationalized architecture to avoid duplication and improve operational efficiency
- maximize system availability and risk management;
- streamline integration and/or interoperability with other relevant Geoscience data systems;
- support and maintain the same level of data quality;
- support continued community involvement in the operation and enhancement of data systems.
GfG Design & Development Principles

- **Broad Application**
  - Design as generic as possible to serve the widest range of applications
  - Use ‘service oriented’ approach to facilitate development of re-usable components
  - Use recognized standards to facilitate interoperability & use by relevant communities.

- **Maximum Utility**
  - Provide valuable services & new opportunities for scientific advance to the community that do not currently exist

- **Sustainability**
  - Maintain & operate systems in professional IT environment that offers highest level of data management standards, access security, risk & contingency management

- **User Friendliness**
  - Make use of systems as easy as possible for any visitor.

- **Science is the driver**
  - Design and implementation of the systems driven by the advancement of science.

- **Community Guidance**
  - Involve the broad science community in design, implementation, and operation of the system to ensure that the systems serve the needs of the community and maintain buy-in and support by the community
GfG: System Operation Principles

- Dynamic modus operandi
  - respond to continuous and expanding demands due to, e.g., changes in community needs, technologies, metadata and interface standards, data types, and policies and procedures for data publication and data access;
  - build linkages with existing and new data systems to create opportunities for new discoveries;

- Community Guidance
  - Any changes and improvements to the systems will be performed and prioritized based on
    - user feedback,
    - advice from the Advisory Committee,
    - technical requirements.
Program Requirements & Status

- Team of highly trained professionals
  - Wide range of expertise
    - scientific proficiency with insight into data generation and data application
    - information technology skills: system engineering, database development, application programming
    - long-term data stewardship
    - project management
  - Capability to bridge between and integrate disciplinary knowledge and information technology expertise
- Level of funding that supports the expert team
Program Requirements

▷ Reliable data management & IT environment
  ➕ Formal Risk and Contingency Management Plans in compliance with Federal and NASA standards
  ➕ Long-term archiving plan based on OAIS approach established between LTA Board and Columbia University Library

▷ Broad involvement of community
  ➕ Collaborations
    • As part of projects (Harvard, OSU, BU, Boise, Kansas)
    • Ad-hoc (GEON, NGDC, CoreWall, etc.)
  ➕ Advisory Committee(s)
  ➕ Wide range of outreach activities
**Tasks**

<table>
<thead>
<tr>
<th>Work Task</th>
<th>Staff</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Management</strong></td>
<td></td>
</tr>
<tr>
<td>Program Management &amp; Directive</td>
<td>Lehnert</td>
</tr>
<tr>
<td>Program Coordination</td>
<td>■ Lehnert, Lenhardt, Vinayagamoorthy, Block</td>
</tr>
<tr>
<td>Project Management</td>
<td>■ Djapic, Block</td>
</tr>
<tr>
<td><strong>System Engineering &amp; Development</strong></td>
<td></td>
</tr>
<tr>
<td>Database design, development, and review</td>
<td>■ Djapic, Arko</td>
</tr>
<tr>
<td>Web site development</td>
<td>■ Gerard</td>
</tr>
<tr>
<td>Search interface development &amp; deployment</td>
<td>■ Fishman, DeGruccio</td>
</tr>
<tr>
<td>Interoperability development (web services, controlled)</td>
<td>■ DeGruccio, Fishman</td>
</tr>
<tr>
<td>Map Interfaces</td>
<td>■ Fishman, Melkonian</td>
</tr>
<tr>
<td>Data loading procedures/scripts</td>
<td>■ Fishman, Djapic</td>
</tr>
<tr>
<td><strong>System Operation</strong></td>
<td></td>
</tr>
<tr>
<td>Database administration</td>
<td>■ Djapic</td>
</tr>
<tr>
<td>Data ingest</td>
<td>■ Fishman, Djapic</td>
</tr>
<tr>
<td>System maintenance, security, backups</td>
<td>■ Strom</td>
</tr>
<tr>
<td>Web site maintenance</td>
<td>■ Gerard, Block</td>
</tr>
<tr>
<td><strong>Data Development</strong></td>
<td></td>
</tr>
<tr>
<td>Data compilation/solicitation</td>
<td>■ Johansson, Block</td>
</tr>
<tr>
<td>Data and metadata entry &amp; quality control</td>
<td>■ Johansson, Block</td>
</tr>
<tr>
<td>User support</td>
<td>■ Johansson, Block</td>
</tr>
<tr>
<td>Data stewardship (long-term archiving, metadata documentation)</td>
<td>■ Lenhardt, Scialdone, Downs</td>
</tr>
<tr>
<td><strong>Outreach &amp; Community Relations, Education</strong></td>
<td></td>
</tr>
<tr>
<td>Publications &amp; presentations</td>
<td>■ Lehnhert</td>
</tr>
<tr>
<td>Workshops</td>
<td>■ Lehnhert, Block</td>
</tr>
<tr>
<td>Exhibits</td>
<td>■ Lehnhert, Block</td>
</tr>
<tr>
<td>Education</td>
<td>■ Lehnhert, Block, Vinayagamoorthy</td>
</tr>
</tbody>
</table>

■ = shared activities  
☑ = project-specific tasks  
☒ = partly shared
Management Functions

▷ Program Manager
   ▶ provides the vision, direction, and priorities
   ▶ insures that the program vision is complimentary with evolving community needs [outreach function]

▷ Project Manager(s)
   ▶ Implement & achieve directions
   ▶ Ensure that timelines, milestones, collaborations, and other requirements are met
GfG Organizational Structure

**Education**
- Karin Block
  - Program Manager Education

**Scientific Directive Community Liaison**
- K. Lehnert
  - Program Director
  - Program Manager IS

**External Collaborators**
- S. Goldstein
  - Senior Scientist

**Program Management**
- C. Lenhardt
  - Program Manager (Data Stewardship)
- (J. Scialdone)
  - Metadata Specialist
- (R. Downs)
  - Archive Specialist
- A. Johansson
  - Data Manager
- Rusty Lotti Bond
  - Data Manager
- Karin Block
  - Data Scientist
  - Project Manager

**Systems Operation**
- (H. Bosch)
  - Operations Manager
- D. Strom
  - System Administrator

**Information Services**
- B. Djapic
  - Project Manager
  - Database Developer
- S. Vinayagamoorthy
  - Program Manager IT

**System Engineering & Development**
- A. Fishman
  - Applications Programmer
- NN
  - Applications Programmer
- A. Gerard
  - Web designer
- R. Arko
  - Senior System Engineer
- A. Melkonian
  - Applications Programmer (MGDS)

**Legend**
- LDEO
- CIESIN

(Funding from GfG projects < 0.1 mm
Direct supervision)
<table>
<thead>
<tr>
<th>Risks</th>
<th>Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>▶ Loss of key personnel</td>
<td>▸ Personnel w/ overlapping expertise; Partnerships, MOU</td>
</tr>
<tr>
<td>▶ Results in delays</td>
<td></td>
</tr>
<tr>
<td>▶ Loss or reduction of funding</td>
<td>▸ Diverse funding basis at CIESIN</td>
</tr>
<tr>
<td>▶ Results in reduced services:</td>
<td>(NASA, USGS, NSF, NOAA, World Bank, UNEP)</td>
</tr>
<tr>
<td>▶ user support &gt; development of new</td>
<td></td>
</tr>
<tr>
<td>▶ functionality &gt; addition of new data</td>
<td></td>
</tr>
<tr>
<td>▶ system maintenance</td>
<td></td>
</tr>
<tr>
<td>▶ Damage or distraction of physical</td>
<td>▸ Risk and contingency plans</td>
</tr>
<tr>
<td>facility</td>
<td></td>
</tr>
</tbody>
</table>
Program Monitoring & Assessment

- Annual reports
- Use statistics
- Advisory committee(s)
  - Structure not defined yet - discussion later today.
Community Issues

- Data sharing (and samples)
  - policies, procedures, and best practices for data reporting

- Data stewardship
  - replication and validation of research findings

- Data citation
  - Credit to authors & data producers
Marine Geoscience
Seismic Data Management System

• preserve academic digital seismic data and supporting information
• help investigators share project data and products
• define, unify and ‘publish’ seismic-related metadata

November 2006
Seismic Data System Components

- **Field Seismic Data Center (LDEO)**
  - collect and serve field data from R/V *Ewing*, *Langseth* and other LDEO portable systems

- **Marine Seismic Data Center (UTIG)**
  - serve older (mostly UTIG) field data from 85 cruises
  - collect and serve processed products
MSDC

- collects and creates (as necessary)
  - acquisition and processing info
  - trace geographic positions, line-by-line
  - SEG-Y binary data with modified headers
  - annotated raster images of sections

seismic sections
map
cruise details
publications
seismic acquisition parameters
LDEO cruise page
processing history
line start/end dates and times
available SEG-Y data - Processed
available SEG-Y data - Field
other information available for this cruise


545x97
November 2006

• collects and creates (as necessary)
  - acquisition and processing info
  - trace geographic positions, line-by-line
  - SEG-Y binary data with modified headers
  - annotated raster images of sections
MSDC Operations

- personnel: (commitment of ~7 months for 2006/7)
  - Lisa Gahagan (manager)
  - Kevin Johnson (systems)
  - Marcy Davis (data intake)
- server-side browser
  - Firefox 1+ (solaris, lunix, osx, windows)
  - Netscape 7+ (solaris, osx); Explorer5.5+ (windows)
- mySQL, Apache, PHP, GeoServer, MapBuilder
- web services in development for:
  - XML metadata, WFS/WMS products
- 3-TB raid disk & 18 TB robotic tape
  - controlled with SAM-FS
  - offsite (tar) storage of entire library

November 2006
MSDC
Acquiring Metadata and Data

- Reluctance of some investigators to contribute processed data
- We’ve reacted to these concerns by:
  - allowing access restriction by data providers
    - we act as broker when user seeks such data
      - usually soon made open access
    - ultimate arbitrator is project funding agency
  - provide citation information with downloads
  - share usage information with data providers
# MSDC Intake Oct 1-Sep 30

<table>
<thead>
<tr>
<th>Cruise</th>
<th>Region/Project</th>
<th>Data Provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>EW0410</td>
<td>Blanco Transform</td>
<td>G. Christeson</td>
</tr>
<tr>
<td>EW0305</td>
<td>Hess Deep</td>
<td>G. Christeson</td>
</tr>
<tr>
<td>EW0210•</td>
<td>Gulf of California (shot)</td>
<td>D. Lizzarralde</td>
</tr>
<tr>
<td>EW0007</td>
<td>Newfoundland Margin SCREECH (partial)</td>
<td>H. Lau</td>
</tr>
<tr>
<td>EW0007•</td>
<td>Newfoundland Margin SCREECH (partial-shot)</td>
<td>S. Holbrook</td>
</tr>
<tr>
<td>EW9705</td>
<td>Iberia Seismic Experiment</td>
<td>D. Sawyer</td>
</tr>
<tr>
<td>EW9705</td>
<td>Iberia Seismic Experiment</td>
<td>Sawyer and Nakamura</td>
</tr>
<tr>
<td>EW9513</td>
<td>Pacific-Australia-Antarctic Plate Circuit (4-chan)</td>
<td>Stock and T. Meckel</td>
</tr>
<tr>
<td>EW9504</td>
<td>California current</td>
<td>M. Lyle</td>
</tr>
<tr>
<td>EW9501</td>
<td>Caribbean Basins</td>
<td>N. Grindlay</td>
</tr>
<tr>
<td>EW9415</td>
<td>LARSE '94</td>
<td>D. Okaya</td>
</tr>
<tr>
<td>NEM003MV</td>
<td>Ice-Age and Millennial-Scale Pacific Intermediate Waters</td>
<td>M. Lyle</td>
</tr>
<tr>
<td>IM0005</td>
<td>Late Holocene Fault Activity in Western Puerto Rico</td>
<td>N. Grindlay, Abrams, and Mann</td>
</tr>
<tr>
<td>VNTR01TW</td>
<td>Eastern Equatorial Pacific Venture Leg 1</td>
<td>SIO data center</td>
</tr>
<tr>
<td>SK97</td>
<td>Sea of Marmara</td>
<td>IRIS data center</td>
</tr>
<tr>
<td>RC2901</td>
<td>Chile Margin</td>
<td>J. Alsop</td>
</tr>
<tr>
<td>RC2606</td>
<td>DSDP Hole 504B</td>
<td>J. Alsop</td>
</tr>
<tr>
<td>RC2604</td>
<td>Barbados Ridge</td>
<td>J. Alsop</td>
</tr>
<tr>
<td>RC2311</td>
<td>Blake Plateau</td>
<td>J. Alsop</td>
</tr>
<tr>
<td>RC2308</td>
<td>Hawaiian Ridge</td>
<td>J. Alsop</td>
</tr>
</tbody>
</table>

( • = includes Ewing shot data to be transferred to LDEO FSDC )

November 2006
# MSDC Cumulative Contents

<table>
<thead>
<tr>
<th>Year</th>
<th>Cruises</th>
<th>Processed SEGY Files</th>
<th>Field SEGY Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>30*</td>
<td>1,193</td>
<td>14,697</td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td>4,798</td>
<td>16,028</td>
</tr>
<tr>
<td>2003</td>
<td>104*</td>
<td>5,171</td>
<td>16,049</td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td>6,007</td>
<td>16,148</td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td>6,966</td>
<td>16,149</td>
</tr>
<tr>
<td>Oct 2006</td>
<td>160</td>
<td>7,942</td>
<td>17,068</td>
</tr>
</tbody>
</table>

*estimate*
# MSDC Yearly Activity

<table>
<thead>
<tr>
<th>Year</th>
<th>Hits (K)</th>
<th>New registered users</th>
<th>File transfers*</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>89</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2002</td>
<td>190•</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2003</td>
<td>271</td>
<td>130</td>
<td>1,116</td>
</tr>
<tr>
<td>2004</td>
<td>378•</td>
<td>116</td>
<td>2,388</td>
</tr>
<tr>
<td>2005</td>
<td>562</td>
<td>138</td>
<td>4,507</td>
</tr>
<tr>
<td>to Oct 2006</td>
<td>1082</td>
<td>133 (+19)</td>
<td>3,205</td>
</tr>
</tbody>
</table>

*=SEGY, navigation and large gifs; + = not monitored • = interpolated gaps

November 2006
**MSDC Registration**

- **Simplified report to AC**
  - no reliable statistics for teaching, research, ‘study’

- **Countries (60):**
  - 40% U.S. (244 of 606)
  - UK (29), India (27), Germany (23), France, Mexico (21), Canada, China (20), Spain, Brazil, Australia (17)

- **Enterprise:**
  - 69% Academic & Government (417)
  - 22% Commercial (132)
  - 6% Personal (35)
  - 4% Undeclared (22)

November 2006
MSDC Tool Set

- view and select SEG-Y binary data, images, navigation
  - map-based search
  - metadata-based search
- external access to metadata:
  - xml metadata service
  - web services support
  - (www.ig.utexas.edu:8080/geoserver/w)
- automated download cart
- custom seismic images

November 2006
Some Digital Marine Seismic Data

- U.S. archive/data centers
  - NSF (LDEO & UTIG) ~200 projects
  - USGS ~100 growing to > 800
  - NGDC ~50 ?
  - SIOSEIS ~50
  - SioExplorer ~few
  - IRIS ~25 marine (more terrestrial)
- Other established archive/data centers
  - Antarctic seismic library ~100 projects
  - IODP (SIO) data bank, ODP CDs ~100 projects
  - Japanese ~20 projects (one of two data centers)
  - France (CNRS, IFEMER) significant
- Un-archived seismic data (chirp, small portable systems)

November 2006
The Promise: Simplifying Discovery
Interoperability

- **objective**—create comprehensive catalogs for users
  - simplifying search and delivery
  - reduces redundant data entry and related errors
  - allows metadata sharing in new ways

- **challenge**
  - in the U.S. are 6+ “archives” with seismic data, and not all data streams are being archived (e.g., chirps, non-UNOLS)
  - difficult to integrate into specialized user interfaces (e.g., geomapapp or MSDC map search, or even Google Earth)
  - data and its relationship to original contributor may become more obscure
  - classes of data access (unrestricted, limited, restricted) limit some usage by higher-level interfaces

November 2006
MSDC Year 4 Objectives

- continue intake of ~40 LDEO cruises and individual scientists
- continue intake from scientists
- create a mechanism for verifying data sharing
- with FSDC/MGDS
  - complete Web Services incoming/outgoing
  - complete interoperability of cruise-level and seismic metadata
- work towards stable funding to:
  - increase efforts in metadata/data collection
  - for participation in design discussions with MMI and LDEO
  - work with other data archives/centers to improve interoperability and minimum metadata requirements for useful interop for seismic data

November 2006
| AntarcticMBS | Documentation/validation and repository for multibeam and other geophysical data from Palmer & Gould  
Recover Legacy data for southern oceans - Core Data, SCS, Deep Sea Photos |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SRDMS (UTIG lead) | Documentation/validation and repository for Multi-Channel Seismic Field Data from Ewing & Langseth  
Processed MCS data from LDEO archives - to UTIG |
| R2K DMS | Integrated data access system for all data collected as part of these programs  
Metadata Catalog for all programs and data repository for relevant legacy, derived and field (orphan) data sets |
|---------|-------------------------------------------------------------------------------------------------------------------|
| MARGINS DMS | Global Bathymetry Synthesis: Build dynamic synthesis of multibeam bathymetry for global oceans - MB Quality Assurance metrics, processing center for Palmer and Langseth MB, custom maps&grids on the fly  
Metadata/Data Catalog for MG&G: Focus on readily recoverable legacy data, work with operators to develop system for modern expeditions |

November 8-9, 2006
Science Support:  
Suzanne Carbotte  
Bill Ryan  
Kerstin Lehnert  
John Diebold  
Admin Support:  
Felicia Taylor  

Technical Team:  
Joyce Alsop  
Robert Arko  
Dale Chayes  
Vicki Ferrini  
Andrew Goodwillie  
Andrew Melkonian  
Suzanne O’Hara  
Rose Anne Weissel
MGDS Partners

Tom Shipley (UTIG) - SeismicReflection
Tim Shank (WHOI) - Ridge2000 Biology
David Becker (TAMU) - MARGINS - ODP
Chris Jenkins (INSTAR) - MARGINS - ODP/DSDP
Diverse Data

- Multibeam Sonar Field Data, Grids and Images
- Multi-Channel and Single Channel Seismic Data
- Temperature Probe Data
- Side-scan sonar imagery
- Magnetic and Gravity Grids
- Magnetotelluric/Electromagnetic Camera Transects
- CTD and MAPR data (conductivity, temperature, transmissivity)
- Biological Species Lists
- Rock, Fluid, Biological Sampling Information
- Instrument Deployment Information
## Access Services

### Data Link

![Data Link](image)

### MGDS: Data Link

**Select Data By:** By Data Type

- **Device Type:** AODC

**Field**
- Program ID:
- Include submersible data
- Investigator:
- Institution:
- Funding Agency:
- Funding Initiative:
- Platform:
- Operator:
- Port Visit:

**Start Date:** 1990-01-01
**End Date:** 2020-12-31

**View Deployed Instruments**

### MARINE GEOSCIENCE DATA SYSTEM
Access Services

GeoMapApp

MARINE GEO SCIENCE DATA SYSTEM
Accomplishments Past 12 months

Data Content

- Lots of new data added - 20,000 new files and doubled total archive volume to 4.5 TB
- Direct data transfer from RPS for Palmer & Gould
- Multibeam QA Working Group meeting
- Cruise metadata from Thompson - Sailing orders from UW

November 8-9, 2006
Accomplishments Past 12 months

Access Services
- Extended Data Link Search Capability
  Search by Data Type and Device Type
- GMA development (Expanded Data Content - SCS, MGT, Help resources)
- GMA to Sea
- Global Bathymetry for GoogleEarth
- Server side Map display (MapServer)
Accomplishments Past 12 months (cont.)

System Operations

- Project management- PEP
- Streamlined Data Ingestion
- Controlled Vocabularies, Web Inventory of Contributions
- Refined metadata forms + land package
- Download acknowledgements
- Password protected access
- Automated Tracking of Growth in Data Holdings
- Improved Tracking of System Use
Interoperability:
• Serving FGDC MB datasets through public interface for NGDC harvesting
• Web Service for controlled vocabularies and data collection
• IRIS - UNAVCO geoWS project

Infrastructure:
• SDSC Data Central
Accomplishments (cont.)

Outreach

- Booth at AGU
- Newsletter Updates & Pamphlet
- Meetings

Meetings & Reports

<table>
<thead>
<tr>
<th>When</th>
<th>Where</th>
<th>What</th>
<th>Who</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec. 11-15, 2006</td>
<td>San Francisco, CA</td>
<td>AGU 2006 Fall Meeting Booth</td>
<td>(all)</td>
</tr>
<tr>
<td>Nov. 8-9, 2006</td>
<td>Palisades, NY</td>
<td>MGDS Advisory Committee</td>
<td>(all)</td>
</tr>
<tr>
<td>Nov. 2-3, 2006</td>
<td>La Jolla, CA</td>
<td>Ridge2000 Steering Committee</td>
<td>Carbotte</td>
</tr>
<tr>
<td>Oct. 31-Nov. 1, 2006</td>
<td>Washington, DC</td>
<td>ORION Cyberinfrastructure Meeting</td>
<td>Carbotte</td>
</tr>
<tr>
<td>Oct. 28, 2006</td>
<td>Palisades, NY</td>
<td>Earth2Class Education workshop</td>
<td>Goodwillie</td>
</tr>
<tr>
<td>Oct. 3-4, 2006</td>
<td>Washington, DC</td>
<td>ORION Cyberinfrastructure Workshop</td>
<td>Chayes, Arko</td>
</tr>
<tr>
<td>Sept. 7-8, 2006</td>
<td>Washington, DC</td>
<td>MARGINS Steering Committee Meeting</td>
<td>Carbotte</td>
</tr>
<tr>
<td>Aug. 15-16, 2006</td>
<td>Seattle, WA</td>
<td>IRIS Web Services/SoA Planning Meeting</td>
<td>Goodwillie</td>
</tr>
<tr>
<td>Jul. 24-26, 2006</td>
<td>Seattle, WA</td>
<td>R2k Endeavour ISS Meeting</td>
<td>Chayes, Arko</td>
</tr>
<tr>
<td>June 21-23, 2006</td>
<td>Woods Hole, MA</td>
<td>QARTOD IV</td>
<td>Carbotte</td>
</tr>
<tr>
<td>June 14-20, 2006</td>
<td>Bremerhaven, Germany</td>
<td>GEBCO meeting</td>
<td>Ryan</td>
</tr>
<tr>
<td>May 15-17, 2006</td>
<td>Tucson, AZ</td>
<td>DLESE Data Services Workshop</td>
<td>Goodwillie</td>
</tr>
<tr>
<td>May 10-12, 2006</td>
<td>Reston, VA</td>
<td>Geoinformatics 2006</td>
<td>Arko, Lehnert</td>
</tr>
<tr>
<td>May 8-10, 2006</td>
<td>Washington, D.C.</td>
<td>Corewall Workshop (IOI)</td>
<td>Ryan</td>
</tr>
<tr>
<td>Apr. 10-12, 2006</td>
<td>Palisades, NY</td>
<td>R2k EPR Field Planning Meeting</td>
<td>(all)</td>
</tr>
<tr>
<td>Mar. 13-14, 2006</td>
<td>Palisades, NY</td>
<td>NSF Site Visit</td>
<td>(all)</td>
</tr>
<tr>
<td>Mar. 10, 2006</td>
<td>Salt Lake City, UT</td>
<td>IRIS/DMS Meeting</td>
<td>Carbotte</td>
</tr>
<tr>
<td>Feb. 18-19, 2006</td>
<td>Honolulu, HI</td>
<td>MARGINS Steering Committee Meeting</td>
<td>Goodwillie</td>
</tr>
</tbody>
</table>
Breakdown of Unique GMA Users by Domain Type (from 12/01/02 to 10/01/06)

Growth in Users of MGDS
<table>
<thead>
<tr>
<th>Domain name - EDU only</th>
<th>Domain name - Non-US Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>butler213mac04.atg.columbia.edu</td>
<td>bridge.earth.ox.ac.uk</td>
</tr>
<tr>
<td>CE3-O-UNET.unet.maine.edu</td>
<td>corp-200-105-245-88-uio.punto.net.ec</td>
</tr>
<tr>
<td>host108-187.lsdepts.ucsb.edu</td>
<td>pavuvu.geo.vu.nl</td>
</tr>
<tr>
<td>deimos.phy.nau.edu</td>
<td>mailgate.igiltd.co.uk</td>
</tr>
<tr>
<td>ki.geosc.psu.edu</td>
<td>earth.crust.irk.ru</td>
</tr>
<tr>
<td>hc652154c.dhcp vt.edu</td>
<td>pc6.gl.rhul.ac.uk</td>
</tr>
<tr>
<td>235-dhcp190.nmc.edu</td>
<td>solitude.delfi.lt</td>
</tr>
<tr>
<td>josephine.whoi.edu</td>
<td>GEOLM15.kfunigraz.ac.at</td>
</tr>
<tr>
<td>dyn-233-152.dyn.columbia.edu</td>
<td>pclab22.earthsci.unimelb.edu.au</td>
</tr>
<tr>
<td>wst100095.terc.edu</td>
<td>axial.ifm-geomar.de</td>
</tr>
<tr>
<td>152-20-21-168.rev.uncw.edu</td>
<td>218-214-37-194.people.net.au</td>
</tr>
<tr>
<td>host108-197.lsdepts.ucsb.edu</td>
<td>adsl9-211.simnet.is</td>
</tr>
<tr>
<td>192-dhcp192.nmc.edu</td>
<td>ns.imr.no</td>
</tr>
<tr>
<td>themountainaingoats.ucsd.edu</td>
<td>behemoth.ifm-geomar.de</td>
</tr>
<tr>
<td>webvpn.ben.edu</td>
<td>bentaiga.ifm-geomar.de</td>
</tr>
<tr>
<td>chickamauga.meas.ncsu.edu</td>
<td>cce-proxy2.dsi.ru</td>
</tr>
<tr>
<td>128-193-142-110.public.oregonstate.edu</td>
<td>ofgsgw3.ori.u-tokyo.ac.jp</td>
</tr>
<tr>
<td>faculty4-843.unh.edu</td>
<td>nat-tor.aster.pl</td>
</tr>
<tr>
<td>nat01-hill-ext.Rutgers.EDU</td>
<td>227iuem22.univ-brest.fr</td>
</tr>
<tr>
<td>ip117-156.pc.jmu.edu</td>
<td>dsl-220-235-152-</td>
</tr>
</tbody>
</table>

Unique Users
Launching GMA in September 2006
Data Volumes Downloaded

Beginning Jan 2004

November 8-9, 2006
Usage Statistics for Marine Geoscience Data System

Summary Period: Last 12 Months
Generated 29-Oct-2006 19:28 EST

Pre-AGU
Upcoming Year

- GoogleEarth for Margins FS and R2K ISS
- Geochemistry Module for GeoMapApp
- Register as OBIS Data provider
- MARGINS Land Data
- International Data Exchange and GEBCO Wksp

- Continue to solicit&ingest data and metadata
Public Dialog

Meetings & Reports

Meetings

<table>
<thead>
<tr>
<th>When</th>
<th>Where</th>
<th>What</th>
<th>Who</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec. 11-15, 2006</td>
<td>San Francisco, CA</td>
<td>AGU 2006 Fall Meeting Booth</td>
<td>(all)</td>
</tr>
<tr>
<td>Nov. 8-9, 2006</td>
<td>Palisades, NY</td>
<td>MGDS Advisory Committee</td>
<td>(all)</td>
</tr>
<tr>
<td>Nov. 2-3, 2006</td>
<td>La Jolla, CA</td>
<td>Ridge2000 Steering Committee</td>
<td>Caroctte</td>
</tr>
<tr>
<td>Oct. 31-Nov. 1, 2006</td>
<td>Washington, DC</td>
<td>ORION Cyberinfrastructure Meeting</td>
<td>Caroctte</td>
</tr>
<tr>
<td>Oct. 28, 2006</td>
<td>Palisades, NY</td>
<td>Earth2Class Education workshop</td>
<td>Goodwille</td>
</tr>
<tr>
<td>Oct. 3-4, 2006</td>
<td>Washington, DC</td>
<td>ORION Cyberinfrastructure Workshop</td>
<td>Chayes, Arko</td>
</tr>
<tr>
<td>Sept. 7-8, 2006</td>
<td>Washington, DC</td>
<td>MARGINS Steering Committee Meeting</td>
<td>Goodwille</td>
</tr>
<tr>
<td>Aug. 15-16, 2006</td>
<td>Seattle, WA</td>
<td>IRIS Web Services/ESOP Planning Meeting</td>
<td>Caroctte, Arko</td>
</tr>
<tr>
<td>Jul. 24-25, 2006</td>
<td>Seattle, WA</td>
<td>R2k Endeavour ISS Meeting</td>
<td>Caroctte</td>
</tr>
<tr>
<td>June 21-23, 2006</td>
<td>Woods Hole, MA</td>
<td>QARTOD IV</td>
<td>Arko, Chayes</td>
</tr>
<tr>
<td>June 14-20, 2006</td>
<td>Bremerhaven, Germany</td>
<td>GEBCO meeting</td>
<td>Ryan</td>
</tr>
<tr>
<td>May 15-17, 2006</td>
<td>Tucson, AZ</td>
<td>DLESE Data Services Workshop</td>
<td>Goodwille</td>
</tr>
<tr>
<td>May 10-12, 2006</td>
<td>Roston, VA</td>
<td>Geoinformatics 2006</td>
<td>Arko, Lehner</td>
</tr>
<tr>
<td>May 8-10, 2006</td>
<td>Washington, D.C.</td>
<td>Corewall Workshop (CJ)</td>
<td>Ryan</td>
</tr>
<tr>
<td>Apr. 10-12, 2006</td>
<td>Palisades, NY</td>
<td>R2k EPR Field Planning Meeting</td>
<td>(all)</td>
</tr>
<tr>
<td>Mar. 13-14, 2006</td>
<td>Palisades, NY</td>
<td>NSF Site Visit</td>
<td>(all)</td>
</tr>
</tbody>
</table>
Data Systems for Science

- Data systems are an enabling resource for obtaining new knowledge
- Enables testing and verification of observations and hypotheses
- Provide access to underutilized data sets for new applications
- Grow the scientific community of interest
- Resources which can be exploited for education and outreach
- Promise of CI requires existence of backbone databases
MGDS – New Content

8th November 2006

Andrew Goodwillie (R2k/MARGINS)
Suzanne O’Hara (AMBS/SDMS)
Overview of Ingestion Procedures

• Pre-cruise contact, metadata forms
• Completed forms/data sent back
• Forms/data checked
• Logged into inventory (reliable back-ups made)
• Streamlined ingestion process:
  • 21 controlled vocabularies
  • 191 foreign key constraints
• PI approval
Example of MGDS Inventory

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>312</td>
<td>email</td>
<td>2006-08-24</td>
<td>arko</td>
<td>2006-08-29</td>
<td>arko</td>
</tr>
<tr>
<td>311</td>
<td>cdrom</td>
<td>2006-08-24</td>
<td>arko</td>
<td>2006-10-24</td>
<td>arko</td>
</tr>
<tr>
<td>310</td>
<td>email</td>
<td>2006-08-23</td>
<td>arko</td>
<td>2006-08-29</td>
<td>arko</td>
</tr>
<tr>
<td>309</td>
<td>email/ftp</td>
<td>2006-08-23</td>
<td>arko</td>
<td>2006-08-28</td>
<td>arko</td>
</tr>
<tr>
<td>308</td>
<td>email</td>
<td>2006-08-16</td>
<td>arko</td>
<td>2006-10-24</td>
<td>arko</td>
</tr>
<tr>
<td>307</td>
<td>ftp</td>
<td>2006-08-18</td>
<td>arko</td>
<td>2006-08-25</td>
<td>arko</td>
</tr>
<tr>
<td>306</td>
<td>email</td>
<td>2006-08-17</td>
<td>arko</td>
<td>2006-08-17</td>
<td>arko</td>
</tr>
<tr>
<td>305</td>
<td>cdrom</td>
<td>2006-08-15</td>
<td>arko</td>
<td>2006-10-17</td>
<td>arko</td>
</tr>
<tr>
<td>304</td>
<td>email</td>
<td>2006-08-14</td>
<td>arko</td>
<td>2006-08-14</td>
<td>arko</td>
</tr>
<tr>
<td>303</td>
<td>email</td>
<td>2006-08-10</td>
<td>arko</td>
<td>2006-09-17</td>
<td>arko</td>
</tr>
<tr>
<td>302</td>
<td>ftp</td>
<td>2006-08-07</td>
<td>arko</td>
<td></td>
<td></td>
</tr>
<tr>
<td>301</td>
<td>cds/dvd</td>
<td>2006-07-29</td>
<td>carbotte</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>ftp</td>
<td>2006-07-25</td>
<td>arko</td>
<td></td>
<td></td>
</tr>
<tr>
<td>299</td>
<td>email</td>
<td>2006-07-21</td>
<td>arko</td>
<td>2006-08-21</td>
<td>arko</td>
</tr>
<tr>
<td>297</td>
<td>disk</td>
<td>2006-07-20</td>
<td>arko</td>
<td>2006-07-31</td>
<td>arko</td>
</tr>
<tr>
<td>296</td>
<td>email</td>
<td>2006-07-20</td>
<td>sohara</td>
<td>2006-07-27</td>
<td>sohara</td>
</tr>
</tbody>
</table>

Done
What's new at the Ridge 2000 Data Portal

October 2006

For EPR 9N site, high temperature HOBO probe data and references now available. Also, expanded options provided for viewing transducers (instruments). Tim Shank's biological samples incorporated for EPR cruise.

A basic cruise entry for TN138 (Butterfield) was created from Sailing Order and Washington Marine Operations office.

For the JdF site, information from McDuff's Portal for Endeavour Data (Pf) database. Now available are the following. Basic cruise entries for AT03-5E (Denerley/Tivey), TN114 (Jackson/White), TN115 (McDuff), TN116 (McDuff), science personnel, Mim/ABE/Exp. Details on information, thermal blanket stations, CTD stations and data. Also detailed basic entries for cruises W9506A and W9506A (Wilcock/Purdy), TN063-06.

September 2006

Raw and edited multibeam bathymetry data for Atlantic cruises AT03-5S and JdF Endeavour segment in 2000 now available.

Simrad EM300 multibeam swath data for EPR 9N site collected during the cruise (contributed by White, Haymon et al.).

For Lau Basin cruise TUM05MV (Tivey), ship track navigation and gravity data downloadable.

Gravity/magnetics data files available for EPR leg DANABERR (Constable).

Basic cruise entry created for JdF NURP 2004 cruise MBARI-2883 (Gillham).

August 2006

Earthquake catalog for 2003-2005 from Tolstoy OBS study at EPR site. Access with password.

Password protection option for accessing data within the database is now an option when they contribute data.

What's new at the Margins Data Portal

November 2006

Sediment analyses data and references now available for Central America TicoFlex II leg EW0204 (Fisher/Silver). For Andy Fisher's TicoFlex II cruise (VANCO2MV), heatflow data, sediment analyses and references, and sediment multibeam bathymetry data files were added.

BLIP data now available for PNG legs, VANCO030MV.

October 2006

Science party personnel and references added for IBM Wiens cruise W9404A, W9405A, KY038B.

September 2006

Raw and processed CTD data now available for PNG VANC legs 14, 15, 16, 17, 20, 21, 23, 25, 27, 28, 29, 30. Additionally, ship track navigation files for PNG legs 22, 23, 24, 25, 26 now available. CTD data added for Goman's Gulf of California cruise, BAJA2NH. Core locations for Kuehl's KM0302 New Zealand area leg added.

Gravity/magnetics data files added for PNG VANC leg, for IBM cruises COOG06MV (Freyer) and COOG07MV (Blomme), and for Central America cruise NEM004MV (Dorman).

Basic cruise entries created for Mariana legs X9404A, X9406A, KY038B (Wiens), including bathymetric grids, CTD deployments and location maps, OBS stations, Seismic shot time files, and ship track navigation files.

August 2006

Basic data entries for the 2000 and 2000 Costa Rica SPS campaigns (O'connor and Schwartz) entered, with links to GPS data and digital chart. In the Gulf of California Focus site, sub-bottom profiler data and ship track navigation data for COOG0A1NH (Galley) available. OBS and CTD operations information for BAJA2NH also added. The PNG VANC leg entries updated with list of publications. CHIRP lines information added for New Zealand area cruises KM0302 (Kuehl) and KM0303 (Alexander). Sample descriptions for dredged rocks on IBM cruise COOG07MV (Blomme) added. Alvin bottom photo Frame/Gabler links activated for AT11-28 Central America cruise (Brown).

Legacy single channel seismic project completed. A total of 415 legs of legacy SCS profiles spanning the global ocean can now be viewed in GeoMapApp.
Highlights: Ridge 2000 - MARGINS

- All R2k programs: metadata, much data
- Time-Critical Study cruises (TCS06NH, AT15-06)
- JdF Portal for Endeavour Data (McDuff)

- All MARGINS cruises: metadata, much data
- Suite of 16 PNG cruises
- Start of land programs

Productive communication with PIs:
- use of metadata forms
- increased submissions
Archive Growth - Terabytes

USCG
RODES
MARGINS
SDMS
AMBS
RMBS

Terabytes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>USCG</td>
<td>0.5</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>RODES</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>MARGINS</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>SDMS</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>AMBS</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>RMBS</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Antarctic Multibeam and Geophysical Data Synthesis Ingestion summary

• Identifying data sets for archiving and serving online

• Data arriving directly from ships (will include PI/RPSC completed documentation) and from PIs

• Identifying and acquiring missing LMG/NBP data from previous cruises

Data that existed at the MGDS prior to October 2006 for the Nathaniel B. Palmer has been ingested into the database.
Data Types from NBP & LMG

- **JGOFS** processed data file
- **MGD77** processed data file
- **ADCP** this data set is current served from another database
- **XBT** ascii format data files
- **RVDAS** store and backup raw and processed data, but do not serve it online except where noted
- **Multibeam** processed and raw swath files, ancillary processing files and survey area grids and images
- **Sound Velocity** both derived and measured products
- **CTD** all levels and types of data provided to the database are served in TAR file bundles as documented stations
- **MOCNESS**
- **Seismic** SEGY data, seismic line information and metadata
- **Sidescan** SEGY data, sidescan line information and metadata
- **Subbottom** SEGY data subbottom line information and metadata
- **Technical and Scientific Reports** (Cruise, calibration, multibeam, Scientific summaries, etc)
- **Satellite Images** images used during cruise
  - **Ice**
  - **Weather**
  - **Isotherm**
- **Scientific Images** images used by science party for this cruise
  - **Ice imagery**
  - **Microscope photographs**
  - **ROV video**
  - **other**
- **Personnel List** including the position of each individual (ie watchstander, chief-scientist, student, ET, MPC ...) and organization or grant that funded participation (ECO, RPSC, OPP99-09734, OPP98-15283, ...)
- **Other data and metadata**: Core Logs, Watch Logs, Chemical analysis, etc
N.B. Palmer Data Link Example
User Feedback

I can see [Data Link] becoming a primary tool for us to find things--even if we have the data in other forms, this format is so handy that it is a great tool to keep track of things.

Julia Smith Wellner
University of Houston

Wow, [Data Link] looks great! I wish we had a site like this for ALL of the IVARS cruises, including those on the Polar Star and Sea.

Vernon Asper
Univ. of Southern Mississippi
## Ingestion summary: Multi-Channel Seismic Field Data

Read from 9-track, 3480 or DAT

Resolve navigation problems, missing shots

QC and compile metadata

Upload to Database

Totals now ~ 2.5 TB

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>CHIEF SCIENTIST</th>
<th>START DATE</th>
<th>PORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>EW9009</td>
<td>Miller, Kenneth</td>
<td>11/1/90</td>
<td>Newark,</td>
</tr>
<tr>
<td>EW9101</td>
<td>Austin, James</td>
<td>2/1/91</td>
<td>Punta Arenas,</td>
</tr>
<tr>
<td>EW9103</td>
<td>McNutt, Marcia</td>
<td>5/12/91</td>
<td>Papeete,</td>
</tr>
<tr>
<td>EW9203</td>
<td>Mutter, John</td>
<td>3/17/92</td>
<td>Gladstone,</td>
</tr>
<tr>
<td>EW9409</td>
<td>McGeary, Sue</td>
<td>7/6/94</td>
<td>Dutch Harbor,</td>
</tr>
<tr>
<td>EW9410</td>
<td>Klemperer, Simon</td>
<td>8/6/94</td>
<td>Dutch Harbor,</td>
</tr>
<tr>
<td>EW9412</td>
<td>Diebold, John</td>
<td>9/12/94</td>
<td>Prince Rupert,</td>
</tr>
<tr>
<td>EW9413</td>
<td>Diebold, John</td>
<td>9/24/94</td>
<td>Prince Rupert,</td>
</tr>
<tr>
<td>EW9501</td>
<td>Driscoll, Neal</td>
<td>2/16/95</td>
<td>Tampa,</td>
</tr>
<tr>
<td>EW9503</td>
<td>Carbotte, Suzanne</td>
<td>5/1/95</td>
<td>Manzanillo,</td>
</tr>
<tr>
<td>EW9509</td>
<td>Liu, Char-Shine</td>
<td>8/23/95</td>
<td>Chilung,</td>
</tr>
<tr>
<td>EW9510</td>
<td>Taylor, Brian</td>
<td>9/28/95</td>
<td>Kaohsiung,</td>
</tr>
<tr>
<td>EW9602</td>
<td>McNutt, Marcia</td>
<td>3/23/96</td>
<td>Lyttelton,</td>
</tr>
<tr>
<td>EW9801</td>
<td>Moore, Gregory</td>
<td>1/26/98</td>
<td>Honolulu,</td>
</tr>
<tr>
<td>EW0104</td>
<td>Fisher, Andrew</td>
<td>4/15/01</td>
<td>Balboa,</td>
</tr>
<tr>
<td>EW0114</td>
<td>Cochran, James</td>
<td>12/7/01</td>
<td>Fremantle,</td>
</tr>
<tr>
<td>EW0207</td>
<td>Carbotte, Suzanne</td>
<td>7/8/02</td>
<td>Astoria,</td>
</tr>
<tr>
<td>EW0305</td>
<td>Christeson, Gail</td>
<td>7/3/03</td>
<td>Balboa,</td>
</tr>
<tr>
<td>EW0307</td>
<td>Holbrook, Steven</td>
<td>8/29/03</td>
<td>Bergen,</td>
</tr>
<tr>
<td>EW0310</td>
<td>McNutt, Marcia</td>
<td>11/14/03</td>
<td>St. Georges,</td>
</tr>
<tr>
<td>EW0408</td>
<td>Jaeger, John</td>
<td>8/22/04</td>
<td>Newport,</td>
</tr>
<tr>
<td>EW0410</td>
<td>Christeson, Gail</td>
<td>10/20/04</td>
<td>Astoria,</td>
</tr>
</tbody>
</table>

*MARINE GEO SCIENCE DATA SYSTEM*
Ingestion summary: Multi-Channel Seismic Processed Data

Processed Data sets recovered from tape, QC, assemble metadata and sent to UTIG

<table>
<thead>
<tr>
<th>Project Code</th>
<th>Type</th>
<th>Region/Campaign</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>UL9905</td>
<td>High-res</td>
<td>Gulf of Cal</td>
<td>5/25/99-6/12/99</td>
</tr>
<tr>
<td>CH0698</td>
<td>High-res</td>
<td>Offshore NJ</td>
<td>5/07/98--5/22/98</td>
</tr>
<tr>
<td>W9605B</td>
<td>High-res</td>
<td>Calif Margin</td>
<td>7/17/96-8/1/9</td>
</tr>
<tr>
<td>EW9602</td>
<td>Ewing</td>
<td>Australs</td>
<td>04/12/96-04/27/96</td>
</tr>
<tr>
<td>EW9501</td>
<td>Ewing</td>
<td>Caribbean</td>
<td>02/21/95-03/19/95</td>
</tr>
<tr>
<td>Ew9413</td>
<td>Ewing</td>
<td>Gorda Rise</td>
<td>09/27/94-10/01/94</td>
</tr>
<tr>
<td>EW9410</td>
<td>Ewing</td>
<td>Alaska</td>
<td>08/06/94-09/01/94</td>
</tr>
<tr>
<td>EW9409</td>
<td>Ewing</td>
<td>Bering Sea</td>
<td>07/09/94-08/02/9</td>
</tr>
<tr>
<td>EW9103</td>
<td>Ewing</td>
<td>Marquesas</td>
<td>05/15/91-06/06/91</td>
</tr>
<tr>
<td>EW9101</td>
<td>Ewing</td>
<td>Antarctic AP</td>
<td>02/08/91-02/28/91</td>
</tr>
<tr>
<td>EW9009</td>
<td>Ewing</td>
<td>NJ Margin</td>
<td>11/04/90-11/17/90</td>
</tr>
<tr>
<td>RC3001</td>
<td>Conrad</td>
<td>MAR Kane FZ</td>
<td>01/17/89-01/20/8</td>
</tr>
<tr>
<td>RC2901</td>
<td>Conrad</td>
<td>Chile Trench</td>
<td>01/09/88-01/31/88</td>
</tr>
<tr>
<td>RC2606</td>
<td>Conrad</td>
<td>Lesser Antilles</td>
<td>03/21/85-04/01/85</td>
</tr>
<tr>
<td>RC2604</td>
<td>Conrad</td>
<td>504B</td>
<td>05/09/85-05/16/85</td>
</tr>
<tr>
<td>RC2405</td>
<td>Conrad</td>
<td>Morocco Leg 1</td>
<td>04/30/83-05/14/83</td>
</tr>
<tr>
<td>RC2406</td>
<td>Conrad</td>
<td>Morocco Leg 2</td>
<td>05/18/83-06/02/83</td>
</tr>
<tr>
<td>RC2311</td>
<td>Conrad</td>
<td>Bahama Carb Bank</td>
<td>10/25/82-11/12/82</td>
</tr>
<tr>
<td>RC2308</td>
<td>Conrad</td>
<td>Hawaii</td>
<td>08/15/82-09/06/82</td>
</tr>
<tr>
<td>RC2114</td>
<td>Conrad</td>
<td>Nor. Margin</td>
<td>09/11/78-10/08/78</td>
</tr>
<tr>
<td>RC2113</td>
<td>Conrad</td>
<td>Norwegian Margin</td>
<td>08/09/78-09/06/78</td>
</tr>
<tr>
<td>RC2106</td>
<td>Conrad</td>
<td>Falklands</td>
<td>03/09/78-03/31/78</td>
</tr>
<tr>
<td>V3618</td>
<td>Vema</td>
<td>Somali Basin</td>
<td>11/18/80-12/14/80</td>
</tr>
</tbody>
</table>
Multibeam Quality Assessment

- Populated \textit{mbinfo} statistics for all 191,138 multibeam files in our holdings
- Have initial quality metric code to the MB-System utility \textit{mbinfo}
- Currently mining the statistics to derive robust \textit{Quality metrics}
Holdings by MB file type

Note: file duration is very different for different sonars
Metrics added to MB-System

dale% mbinfo -q -x -f 41 -i raw/sb20030860000.mb41

< Mbsystem >
  < Mbinfo >
    < SwathFileName >
      sb20030860000.mb41
    </ SwathFileName >
    < MbioDataFormatId >
      41
    </ MbioDataFormatId >
    < Mbqam >
      < Version >
        1.0
      </ Version >
      < Drops >
        0.386190
      </ Drops >
    </ Mbqam >
  </ Mbinfo >
</ Mbsystem >
Dropped Beams per file by type

Note: Different sonars and instances behave differently!
Progress & Plans for Our Data Access Tools

• Data Link
• GeoMapApp
• Google Earth

Advisory Committee Meeting

November 8-9, 2006
Search by Data Type

List is dynamically generated from database
for GeoMapApp

• Reduce Effort to Add New Functionality

• Concentrate Effort to Add More Content
1) Data tables
For example, the LDEO piston core collection.
Deep Sea Core Repository

Lamont-Doherty Earth Observatory of Columbia University

Visual Core Description for VM19-67

Date described: 14 June 1966. Described by: R. Baker. 0-20 cm: Lutite, moderate to dark yellowish-brown (10 YR 5/3). Semi-moist and compact; broken and dry in the top 5 cm. Burrow mottled and homogeneous in texture. The lutite is slightly sandy, and contains approximately 40% carbonate (mainly Globigerina, Globorotalia), little silt, and an occasional manganese micronodules. Gradational, intermottled bottom contact. 20-227 cm: Lutite, dusky brown (5YR2/2), moist, uniform and featureless, mottled with the overlying lutite in the upper 30 cm. Homogeneous in color and texture. Broken by regularly spaced cracks approximately 5-10 cm. apart due to slight drying of the surface. Carbonate content is reduced to 10-15%, and manganese micronodules more prevalent (accounting for the dark color of the lutite). Mottled, and gradational bottom contact. 227-270 cm: Lutite, moderate yellowish-brown (10YR5/4). Burrow mottled by the organic reworking of the overlying lutite. Similar to the overlying lutite, but higher in carbonate content. Gradational bottom contact. 270-320 cm: Lutite, moderate brown (5YR3/4), faintly mottled by the overlying lutite, otherwise similar in composition and texture. Mottled and gradational bottom contact. 320-442 cm: Calciolutite, dark yellowish-orange (10YR6/6). Burrow mottled and grading to moderate brown (5YR3/4) at 370 and 430 cm. Little silt content; carbonate approximately 60%, (Globigerina and Globorotalia). Few manganese micronodules, shell fragments and sponge spicules. The bottom 8 cm. is a dried, stained indurated fragment. Note: Iron oxidation stains are common along most of the core. Trigger Weight Date described: 23 February 1970. Described by: J. Daubenspeck. 0-16 cm: Foraminiferal chalk oozes, between grayish orange pink and Pale yellowish brown (5 YR 7/2 and 10 YR 6/2), dry, hard and fragmented. Carbonate content high. Coarse fraction 35-40%, containing abundant foraminifera fragments, rare echinoid spines and fish teeth. Inorganic elements negligible. Basal contact gradational in color and composition. 16-24 cm: Clay, dark yellowish brown (10 YR 4/2), dry, hard and fragmented. Carbonate content low. Coarse fraction about 5%, containing common foraminifera fragments, fish teeth and bone fragments. Inorganic elements include common phillipsite, rare palagonite, quartz and manganese.

Funded by the National Science Foundation
### Lamont-Doherty Photo Station Locations

<table>
<thead>
<tr>
<th>Name</th>
<th>Station</th>
<th>Date</th>
<th>Time</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Text_Lat</th>
<th>Text_Lon</th>
<th>Depth_m</th>
<th>Orientation</th>
<th>Fe_m</th>
<th>Tracks&amp;Trails</th>
<th>Fish</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH61-001</td>
<td>18</td>
<td>03-08-66</td>
<td></td>
<td>36.333</td>
<td>-7.400</td>
<td>36°33'N</td>
<td>007°24'W</td>
<td>450</td>
<td>54°40'E</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>RC01-001</td>
<td>05-12-62</td>
<td></td>
<td></td>
<td>27.833</td>
<td>-76.700</td>
<td>27°50'N</td>
<td>076°42'W</td>
<td>733</td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>RC01-002</td>
<td>17-12-62</td>
<td></td>
<td></td>
<td>31.067</td>
<td>-78.450</td>
<td>31°04'N</td>
<td>078°27'W</td>
<td>410</td>
<td></td>
<td>4</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>RC01-003</td>
<td>17-12-62</td>
<td></td>
<td></td>
<td>31.450</td>
<td>-76.550</td>
<td>31°27'N</td>
<td>076°33'W</td>
<td>1555</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>RC01-004</td>
<td>18-12-62</td>
<td></td>
<td></td>
<td>32.100</td>
<td>-74.200</td>
<td>32°06'N</td>
<td>074°12'W</td>
<td>2560</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>RC01-005</td>
<td>20-12-62</td>
<td></td>
<td></td>
<td>31.150</td>
<td>-74.883</td>
<td>31°09'N</td>
<td>074°53'W</td>
<td>1782</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>RC04-003</td>
<td>12-03-63</td>
<td>2053</td>
<td></td>
<td>28.333</td>
<td>-77.533</td>
<td>28°20'W</td>
<td>077°32'W</td>
<td>548</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>RC04-004</td>
<td>13-03-63</td>
<td>2303</td>
<td></td>
<td>29.783</td>
<td>-75.667</td>
<td>29°47'N</td>
<td>075°40'W</td>
<td>677</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>RC04-005</td>
<td>04-03-63</td>
<td>1321</td>
<td></td>
<td>30.633</td>
<td>-76.267</td>
<td>30°38'N</td>
<td>076°18'W</td>
<td>1775</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

**Elevation Data Sources**
Photos with manganese nodules
LDEO nepheloid layer profiles
Concentration of nepheloid layer (ug/cm²)
<table>
<thead>
<tr>
<th>name</th>
<th>longitude</th>
<th>latitude</th>
<th>min_depth</th>
<th>max_depth</th>
<th>date</th>
<th>link</th>
</tr>
</thead>
<tbody>
<tr>
<td>st_0</td>
<td>53.0167</td>
<td>5.0333</td>
<td>-5104.0</td>
<td>-4571.0</td>
<td>31.07.19</td>
<td>[<a href="http://www">http://www</a>]</td>
</tr>
<tr>
<td>st_1</td>
<td>60.575</td>
<td>4.9917</td>
<td>-3317.0</td>
<td>-3301.0</td>
<td>17.04.19</td>
<td>[<a href="http://www">http://www</a>]</td>
</tr>
<tr>
<td>st_2</td>
<td>79.0667</td>
<td>0.9167</td>
<td>-4565.0</td>
<td>-4423.0</td>
<td>31.08.19</td>
<td>[<a href="http://www">http://www</a>]</td>
</tr>
<tr>
<td>st_3</td>
<td>78.975</td>
<td>2.925</td>
<td>-4137.0</td>
<td>-3941.0</td>
<td>29.08.19</td>
<td>[<a href="http://www">http://www</a>]</td>
</tr>
<tr>
<td>st_4</td>
<td>78.0833</td>
<td>6.1167</td>
<td>-3017.0</td>
<td>-2962.0</td>
<td>26.02.19</td>
<td>[<a href="http://www">http://www</a>]</td>
</tr>
<tr>
<td>st_5</td>
<td>88.93</td>
<td>0.1667</td>
<td>-4236.0</td>
<td>-4235.0</td>
<td>24.04.19</td>
<td>[<a href="http://www">http://www</a>]</td>
</tr>
<tr>
<td>st_6</td>
<td>88.5217</td>
<td>1</td>
<td>-4405.0</td>
<td>-4402.0</td>
<td>24.04.19</td>
<td>[<a href="http://www">http://www</a>]</td>
</tr>
<tr>
<td>st_7</td>
<td>89.0157</td>
<td>3.0333</td>
<td>-3574.0</td>
<td>-3189.0</td>
<td>10.09.19</td>
<td>[<a href="http://www">http://www</a>]</td>
</tr>
<tr>
<td>st_8</td>
<td>81.9333</td>
<td>5.95</td>
<td>-3705.0</td>
<td>-3431.0</td>
<td>24.02.19</td>
<td>[<a href="http://www">http://www</a>]</td>
</tr>
<tr>
<td>st_e</td>
<td>63.9667</td>
<td>6.0333</td>
<td>-3840.0</td>
<td>-3843.0</td>
<td>24.02.19</td>
<td>[<a href="http://www">http://www</a>]</td>
</tr>
</tbody>
</table>

- depth
- temp_deg_c
- salinity_ppt
- oxygen_ml/l
- phosphate_microgram_atoms/l
- phosphorous_microgram_atoms
- silicate_microgram_atoms
- nitrite_microgram_atoms
- nitrate_microgram_atoms
Topography, magnetics and gravity profiles
Sediment thickness

Additional gridded datasets
Seafloor ages
GEBCO bathymetry
All LDEO single channel reflection profiles
Single Channel reflection profiles can be downloaded from Data Link.
Looking for big elephants:

• LDEO bottom photo collection (6000 stations)
• IRIS earthquakes
• ALVIN photos
• Greenland & Antarctic Ice Cores
For Google Earth

• Convert from ARC GIS projects in focus areas

• Add terrestrial datasets
Education Component of the Ridge2K component of the Marine Geoscience Data System

Kim Kastens

Presentation to Advisory Committee
8 November 2006
Education Component of the Ridge2K component of the Marine Geoscience Data Management System

- Input from education-oriented users into design of data access and display tools
- Learning Science research
- Development, dissemination, and testing of exemplary data-rich student activities
  - Earth Exploration Toolkit
  - Data Puzzles
  - Conceptests
Input from education users into tool design

• 2005 DLESE Data Services Workshop
• Dedicated workshop on use of Ridge data in education (2005)

• MARGINS Education & Public Outreach workshop
• 2006 DLESE data services workshop
• Earth2Class (K-12 teachers)
• L-DEO summer interns
• Hamilton College undergrads

http://swiki.dlese.org/DataSvcsWkshp-04/140
Input from education users into tool design

Recommendations implemented:
• Map scale (adjusts with zoom)
• Inset location map
• Distance-measuring tool
• Clearer links to information about how data were collected
• Overlay of place names (partial)
• Visual indicator when system is thinking
• User manual or “Help” feature

Compelling recommendations remaining:
• Depth scale for earthquakes
Learning Science Research

*Investigator:* Sandra Swenson, Teachers College doctoral student

*Participants:* 120 total: 8th, 10th, 12th graders

*Questions:*
  - What do you think this is?
  - How do you think this was made?
  - What do you think this is useful for?
What do you think this is?

- geographical map (15)
- world map (20)
- underwater features (13)
- topography (29)
- depth of water “dark area is higher”
- plate boundaries (3)
- geological map (3)
- “shows temperature by different blue colors”
- shows where there are high levels of sodium (for fishing)”
- “a map showing tides”
- world climates (e.g. white near N. pole is snow and glaciers)”
Data-rich Activities for Students

Earth Exploration Toolkit (summer 2005)
• full access to professional quality datasets

Lamont Data Puzzles (summer 2006)
• carefully pre-selected data snippets

Conceptests (forthcoming)
• in-class response/discussion/re-response
Data-rich Activities for Students

Earth Exploration Toolbook example:
• bathymetry & isochrons from GeoMapApp
• by Jeff Thomas, Fairfield High School (CT), Earth Science class

http://serc.carleton.edu/dev/eet/rodes_6/index.html
Lamont Geoscience Data Puzzles

- Low-barrier-to-entry for teachers
- High insight to effort ratio for students
- Offers an “aha” moment
- Built around carefully selected (but authentic) geoscience data
- Time-efficient
- Detailed pedagogical content knowledge (PCK) documentation for teachers

One figure

One question

Another question

To explore further see….
Lamont Geoscience Data Puzzles

“How much heat is released from a hydrothermal vent?”
By Rosemarie Sanders, Ramapo High School.

VS.
Lamont Geoscience Data Puzzles

“Ocean Floor Spreading”
by Greg Hofer, Baccalaureate School for Global Education.

Vs.

[Diagram showing ocean floor spreading patterns on both sides, with observed and mirrored sections.]
• The “aha” moment
• Alignment with science education standards
• Skills & understandings needed
• Step-by-step solution
• Common misconceptions and mistakes
• Tough questions
• To learn more...

Q: If the water is above 100°C, why isn’t it steam?

A: The water at the bottom of the ocean is at high pressure from the weight of the overlying water. The high pressure keeps the water in its liquid state. At 3000m water depth, the boiling point of water is around 410°C.
“Conceptests”

- Question assessing student understanding of principal concepts is projected mid-lecture.
- Students answer, results tallied immediately, students discuss in small groups, students answer again.
- Reported benefits: student engagement, attendance, understanding, teacher pacing

References:
- Question bank at: serc.carleton.edu/introgeo/interactive/ctestexm.html
Example Mid-Ocean Ridge Concepttest

If you dredged a rock sample from the seafloor at the spot labelled “X” on the map, which rock type would you be most likely to recover?

(a)  
(b)  
(c)
Education Component of the Ridge2K component of the Marine Geoscience Data Management System

- Input from education-oriented users into design of data access and display tools
- Learning Science research
- Development, dissemination, and testing of exemplary data-rich student activities
  -- Earth Exploration Toolkit
  -- Data Puzzles
  -- Conceptests

Questions or Suggestions?
“INTEROPERABLE” = TRANSPORTABLE + UNDERSTANDABLE

Q: Why interoperability?

A: To facilitate research -
   • enable interdisciplinary work
   • make data more broadly, easily available to non-specialists
   • let users choose their tools
   • enforce data documentation
Q: How to be interoperable?

A: Service-Oriented Architecture (SOA) - providers satisfy requests from consumers -

- descriptive, not instructive
- standard format, structure, vocabulary ➤ our current focus
- extensible
- discoverable
Q: How to implement SOA?

A: Web Services - Web-based applications that use open, XML-based standards to exchange data with clients -

• Registry services (*catalogs*)

• Access services (*search and retrieval*)

• Ontology services (*concepts*)

• Transform services

• Workflow services
Data Organization

Collection
(count: 1513)

Data Set
(count: 6114)
(85 types)

File
(count: 210,048)
Work in Progress -

1. Catalog services: Collections, Vocabularies
2. Data set service: Multibeam data
3. OGC Web Catalog, Map, and Feature Services
4. Ontology development: Marine Platforms
5. Integration with SDSC and TeraGrid
1. Catalog services
   *(joint work with CIESIN/GfG, UTIG/SDC, BRG/IODP, SIO/GDC)*
   
   • Vocabulary service
     [http://www.marine-geo.org/ws/getVocabulary/?id=data_type](http://www.marine-geo.org/ws/getVocabulary/?id=data_type)
   
   • Collection service
     [http://www.marine-geo.org/ws/getCollection/?id=EW0207](http://www.marine-geo.org/ws/getCollection/?id=EW0207)
   
   • records built live from database, expose entire holdings
1. Catalog services
2. Data set service: Multibeam data  
(joint work with NGDC, CCOM)

- create FGDC metadata records
- validate against NOAA Metadata Manager & Repository (NMMR) schema
- publish in OAI-PMH provider
2. Data set service: Multibeam data
3. OGC Web Catalog/Map/Feature Services
(joint work with UNAVCO, IRIS)

- partner data centers serve OGC-compliant GIS layers
- extend layers to add 3rd (elevation) and 4th (time) dimensions via new WMS-T standard
- integration demo at AGU Fall 2006
3. OGC Web Catalog/Map/Feature Services
4. Ontology development: Marine Platforms
(joint work with Graybeal/Bermudez-MBARI, Lowry-BODC, Raskin-JPL, O’Neill-CCLRC, Drewry-UAH)

- develop and publish ontology for marine platforms
- develop SensorML profile for multibeam sonars
- OOSTethys testbed development *(joint with IOOS + OOI)*
  to model research platforms in OGC SOS
4. Ontology development: Marine Platforms

Technical papers 9:05 - 10:45 am.
- New Architecture for the Sensor Web: the SWAP-Framework
  Deshadran Moodley, Jagga Simons
- ES3N: A Semantic Approach to Data Management in Sensor Networks
  Mike Lewis, Delroy Cameron, Shalhua Xue, Budak Aydinlar
- Semantic Connection between Everyday Objects and a Sensor Network
  Stephen Temple, Yukihiro Hidaka, Satoru Satake, Hideyuki Arakawa
- A Marine Platforms Ontology: Experiences and Lessons
  Luis Bermudez, John Graybeal, Robert Arko

Morning tea 10:45am - 11:05am.
5. Integration with SDSC and TeraGrid

- live mirror of entire holdings including MCS field data
- leverage existing NSF facility
- addresses sustainability + heterogeneity
5. Integration with SDSC and TeraGrid

5.1. SDSC User Services - Data Allocations

- Data Central: Search Allocations > Allocation

**Marine Geoscience Data System (MGDS)**

**URL:** [www.marine-gao.org](http://www.marine-gao.org)

**Field:** Marine Geology and Geophysics

**Project Summary:**
MGDS is a multi-disciplinary repository for marine geoscience data and services for the NSF Ridge 2000, MARGINS, and U.S. Antarctic Programs, representing ~1000 researchers, educators, and students. We currently host ~200,000 data objects from ~1500 field expeditions and lab programs dating from 1952, totalling ~506GB currently served online and growing at a rate of ~100GB/year. Our holdings consist of both original field data and post-field analytical products.

**PI:**
Robert Arko
Lead Programmer/Analyst
Lamont-Doherty Earth Observatory
61 Route 9W
Pallisades, NY 10964
(845) 365-8752
arko@ldeo.columbia.edu

**SDSC Contacts:**
Ken Nunes
(858) 534-5161
kunnus@sdsc.edu
Addressing Political Challenges: Building a National and International Geoinformatics System
Definitions

- **Cyberinfrastructure**: “comprehensive digital environments that become interactive and functionally complete for research communities in terms of people, data, information, tools, and instruments” (Atkins Report)

- **Geoinformatics**: “A distributed, integrated digital information system and working environment that provides innovative means for the study of the Sun-Earth system and other planets through the use of advanced information technologies.”
The Challenge

“The challenge of Cyberinfrastructure is to integrate relevant and often disparate resources to provide a useful, usable, and enabling framework for research and discovery characterized by broad access and “end-to-end” coordination.”

Fran Berman, Director San Diego Supercomputer Center
SBE/CISE Workshop on Cyberinfrastructure for the Social Sciences

Most of the technology exists
Challenges are cultural and organizational

L. Allison, SedDB workshop 2004
Building a Geoinformatics System

- Link & integrate data, information, applications
- Leverage experiences & achievements
- Reduce redundancy

- Communication
- Coordination
- Collaboration
National Efforts

GSA Geoinformatics Division

Geoinformatics Division Officers, 2007

4 Members Chair—1 year; First Vice-Chair—1 year; Second Vice-Chair—1 year;
Secretary-Treasurer—2 years

Chair: A. Krishna Sinha, Virginia Tech, Blacksburg, pitlab@vt.edu
First Vice-Chair: Randy Keller, University of Oklahoma, Norman
Second Vice-Chair: Linda Gundersen, U.S. Geological Survey, Reston
Secretary-Treasurer: Dogan Seber, University of California San Diego, seber@sdsc.edu
Past Chair: not applicable until 2008

GSA Council Liaison Nancy McMillan, New Mexico State University, Las Cruces
Advisory Board Chair: Robert D. Hatcher, UT, Knoxville, bobmap@utk.edu
Mike Williams, University of Massachusetts, Amherst
Chaitan Baru, San Diego Supercomputer Center, University of California San Diego
Rob Raskin, Jet Propulsion Laboratory, Cal Tech, Pasadena
Art Goldstein, Earth Sciences Division, National Science Foundation
Susan Eriksson, Education and Outreach Director, UNAVCO, Boulder

Nominating Committee Chair: Cal Barnes, Texas Tech, Lubbock, cal.barnes@ttu.edu
Ramon Arrowsmith, Arizona State University, Tempe
Leslie Ruppert, U.S. Geological Survey, Reston
National Efforts

- AGU Focus Group “Earth & Space Science Informatics”

Earth and Space Sciences Informatics (ESSI) Focus Group

AGU’s ESSI Focus Group serves to facilitate communications and coordinate activities related to issues of data management and analysis, large-scale computational experimentation and modeling, and the hardware and software infrastructure needs to span the range of scientific topics of interest to the Union.

If you would like to contribute material to ESSI and the ESSI web site contact the web site manager - Tom Narock tom.narock@gsc.nasa.gov
We are interested in all material related to the interests of the group. If your sending figures please include captions.

Affiliate With ESSI

Update your AGU membership record to reflect your interest in ESSI.
Go to Member Login Page and select “Earth and Space Sciences Informatics” as your primary or secondary affiliation

Officers

- ESSI Chair: Jean-Bernard Minster
National Efforts

AGU Town Hall Meeting
“Envisioning the Future of Earth Science Data and Knowledge Access Through a Broad Geoinformatics Collaboration”

Organizers:
Linda Gundersen (USGS), Peter Fox (UCAR), Kerstin Lehnert (Columbia U), Deborah McGuinness (Stanford U), Krishna Sinha (Virginia Tech), Walter Snyder (Boise State U)

EOS Article (Nov 14, 2006): “Towards Broad Community Collaboration in Geoinformatics”
“Towards Broad Community Collaboration in Geoinformatics”

- “Healthy growth in Geoinformatics will require multi-agency and disciplinary society partnerships as much as collaboration among individual projects.
- No single group can proclaim that they are the universal organizer, consolidator, or focal point for Geoinformatics.
- Instead, any Geoinformatics organization that is accepted on a broad and encompassing basis must come from a long and deliberate process of bringing people together where the focus and leadership boils up out of that group.”
Town Hall: Objectives

- Promote communication among GI efforts nationally and internationally, across agencies, disciplines, and organizations
- Involve scientists and educators
- Reach out to the general CI community to tie in related, complementary, and infrastructure efforts
- Seek comments on extending partnerships nationally & internationally
Town Hall: Panel

- Participants
  - Kelvin Droegemeier
  - Randy Keller
  - Jens Klump
  - David Maidment
  - Dave Lambert
  - Tim Grove
  - Larry Smarr
Panel Topics

- What are the challenges in making GI a collaborative national or international effort?
- How do we as geoscientists move the Geoinformatics community closer to a coherent “alliance” of projects and people who truly want to move Geoinformatics forward to fulfil its promises?
- How can funding agencies support this initiative? What role can professional societies play? What funding opportunities currently exist?
Panel Topics

- Are sustainable collaborations between discipline and computer scientists in place now? If so, how do they create balance in a Geoinformatics effort between a research frontier and a service to science? If not, how can this be achieved?
- What fresh, new approaches can help engage the broad science and education community in Geoinformatics?
National Efforts

Coordination Activities and Workshops

NCAR will host a series of workshops to share information on NSF and related funded Cyberinfrastructure initiatives that may provide opportunities for leveraging shared cyberinfrastructure across environmental observatory programs. A series of presentations describe technical innovations, effective management and coordination practices, and other lessons learned in the implementation of successful existing CI projects. Breakout sessions are aimed to provide opportunities for discussion and networking between discipline scientists and CI practitioners.

Next workshop

Announcements of Opportunity

How to Participate

Click Here To Register

Past workshops:

ORION - Ocean Observing Initiative CyberInfrastructure
National Efforts

- ARL meeting - Suzanne could talk about this
International Efforts

Workshop Proposal: International Data Access for Studies of Global Tectonics at Continental Margins and Mid-Ocean Ridges

May 9-11, 2007 Kiel Germany

Co-Sponsors: InterMargins, InterRidge, MARGINS, Ridge2000
Workshop Goals

1. To explore ongoing data management efforts relevant to continental margin and mid-ocean ridge science goals within partner countries. - overview of existing data systems that host data relevant for continental margins and mid-ocean ridge related research and discussion of currently available tools for data access and visualization.

2. To devise a strategy for the sharing and exchange of key data sets of mutual scientific interest for the sponsoring programs. As a first step, identify high priority data sets of mutual interest need. Then issues required to facilitate data exchange need to be defined, including data and metadata standards, and the appropriate database interfaces for interoperability.
Draft Agenda Day 1

- Science presentations on needs of InterRidge/InterMargin scientists for data and visualization tools
  - InterMargins, InterRidge, Ridge2000, MARGINS

- Existing Data Systems Relevant for InterMargins/InterRidge research: What data is currently available and how is it served?
  - ESONET, NeptuneCanada, ChESSbase, GEOROC, JAMSTEC Data Centers, Ridge2000/MARGINS DMS, GfG

- Models of Existing/Developing International Data Systems for Earth Science Data
  - SeaDataNet – European marine data system, FDSN - seismic data, WDC – Bremen, WDC- NGDC - Boulder, SCOR – International cruise data
Draft Agenda Day 2

Emerging Technologies
- Data Registration Services - Unique identifiers for data and samples
- Web Services
- Virtual Globes

Working Group Discussions
- What are the current obstacles to international data exchange for continental margins and mid-ocean ridge research?
- Strategy for interoperability including what data/metadata should be exchanged?
- How do we achieve data contribution to national/institutional archives?
Draft Agenda Day 3

- **Plenary Session**

  Summary of working group discussions and formulation of key recommendations.

  Convenors: Suzanne Carbotte, Kerstin Lehnert (LDEO), Seiji Tsuboi (Jamstec), Willi Weinrebe (GEOMAR)
SCOR Meeting 2006

- Suzanne can talk about this
GEBCO 2006

- Will host upcoming meeting in 2007
National Efforts

Associated Research Libraries

To Stand the Test of Time

Long-term Stewardship of Digital Data Sets in Science and Engineering

A Report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe

September 26–27, 2006
Arlington, VA
This is a report of a two-day workshop that examined the role of research and academic libraries with other partners in the stewardship of scientific and engineering digital data. Workshop participants explored issues concerning the need for new partnerships and collaborations among domain scientists, librarians, and data scientists to better manage digital data collections; necessary infrastructure development to support digital data; and the need for sustainable economic models to support long-term stewardship of scientific and engineering digital data for the nation’s cyberinfrastructure.
- Historically, universities have played a leadership role in the advancement of knowledge and shouldered substantial responsibility for the long-term preservation of knowledge through their university libraries. An expanded role for some research and academic libraries and universities, along with other partners, in digital data stewardship is a topic for critical debate and affirmation.
Recommendations to NSF

- Support pilot projects to develop partnerships between data centers and research libraries
- Support training programs for data scientists – curators
- Support development of improved tools for automatic metadata harvesting
- Require inclusion of data management plans in proposals
International Efforts

Workshop Proposal: International Data Access for Studies of Global Tectonics at Continental Margins and Mid-Ocean Ridges

May 9-11, 2007 Kiel Germany

Co-Sponsors: InterMargins, InterRidge, MARGINS, Ridge2000
Workshop Goals

1. To explore ongoing data management efforts relevant to continental margin and mid-ocean ridge science goals within partner countries. - overview of existing data systems and of currently available tools for data access and visualization.

2. To devise a strategy for the sharing and exchange of key data sets of mutual scientific interest for the sponsoring programs. - identify high priority data sets of mutual interest, issues required to facilitate data exchange need to be defined, including data and metadata standards, and the appropriate database interfaces for interoperability.
Draft Agenda Day 1

- Science presentations on needs of InterRidge/InterMargin scientists for data and visualization tools
  InterMargins, InterRidge, Ridge2000, MARGINS

- Existing Data Systems Relevant for InterMargins/InterRidge research: What data is currently available and how is it served?
  ESONET, NeptuneCanada, ChESSbase, GEOROC, JAMSTEC Data Centers, Ridge2000/MARGINS DMS, GfG. Others?

- Models of Existing/Developing International Data Systems for Earth Science Data
  SeaDataNet – European marine data system, FDSN – seismic data, WDC – Bremen, WDC- NGDC – Boulder, SCOR – International cruise data, Others?
Draft Agenda Day 2

- **Emerging Technologies**
  - Data Registration Services - Unique identifiers for data and samples
  - Web Services
  - Virtual Globes

- **Working Group Discussions**
  - What are the current obstacles to international data exchange for continental margins and mid-ocean ridge research?
  - Strategy for interoperability including what data/metadata should be exchanged?
  - How do we achieve data contribution to national/institutional archives?
Draft Agenda Day 3

- **Plenary Session**
  
  Summary of working group discussions and formulation of key recommendations.

  Convenors: Suzanne Carbotte, Kerstin Lehnert (LDEO), Seiji Tsuboi (JAMSTEC), Willi Weinrebe (GEOMAR)
Additional International Meetings

- SCOR - 2006
- GEBCO 2007
Addressing Cultural Challenges:

Referencing data and data sources-the problem of proper credit
Global geochemical databases (GEOROC, PetDB, NAVDat) have changed the way we do geochemistry

- Need to facilitate seamless transfer of data from publication to database

- Need to be ensure that sample metadata and analytical metadata are satisfactory

- Need to be able to link data taken different times from different labs on same sample
Establish common metadata guidelines for journals that publish significant geochemistry:

Chemical Geology, CMP, EPSL, GCA, G-cubed, Geology, GRL, J Petrology, Lithos, JVGR, JGR, MarChem, Paleoceanography, Nature(?), Science(?), etc.

Unique sample identifier (IGSN)
This may solve the problem of poor sample metadata.

Essential component of successful database - contains sample metadata, allows samples to be followed through its analytical history.

Tracks samples and subsamples.

We should start using it now
Addressing Cultural Challenges:

Referencing data and data sources - the problem of proper credit
Increasing Use of Geochemical Databases

At least 118 citations since 2000.
How exactly do we determine and compare scientists’ contributions to knowledge? ....

Although a large number of different measures have been used occasionally,² the two major quality indicators are based on publication counts or citation counts, respectively.
View of Geochemists:

“If you use data I generated, you should cite my work”

Why is this such a big issue:

1. Ego

2. Evaluations
The National Research Council
Assessment of Research Doctorate Programs, 2006-07

Assessment Overview and Implications for Columbia

Lucy Drotning
Associate Provost for Planning and Institutional Research

John Scanlon
Data Manager, Planning and Institutional Research
Background

Goal

The stated goal of the study is to “help universities improve the quality of these programs through benchmarking; provide potential students and the public with accessible, readily available information on doctoral programs nationwide; and enhance the nation’s overall research capacity.”

The study will consist of:

1. Collection of quantitative data through questionnaires

2. Data on faculty publications, citations, and research activity and student dissertation keywords

3. Faculty opinion on the relative importance of measures of program quality

Who is doing what?

• NRC
• Mathematica Policy Research (MPR)
• Columbia
Increasing Use of Geochemical Databases

Kerstin is happy that there are 118 citations of PetDB. These papers use data but cite the database, often not the papers.
Citations of Databases, not data papers

Figure 5. Mantle tetrahedron of Hart et al. [1992]. Arrays from end-member defining island chains have been plotted using the GEOROC database and data presented in this manuscript.…..
From Hauri et al. EPSL 2006 “Partitioning of water during melting of the Earth's upper mantle…”

Fig. 5. H$_2$O-REE systematics of MORB (dots), OIB (crosses), arc basalts (triangles) and the results of modeling calculations (lines, see text). Data are from the PetDB and GeoRoc databases, and unpublished data (EHH).
Salters and Stracke G-cubed 2004
“Composition of the depleted mantle”

From the text: “the average MORB composition results from compiling MORB from the PetDB database”

Supplementary material: Tables A1 and A2. Table A1 contains the data used to obtain the average MORB composition. This file contains three worksheets. References contains the references for the data given by the PetDB database.
Citations of Databases and data papers

<table>
<thead>
<tr>
<th>Database</th>
<th>Authors</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-cubed</td>
<td>Salters and Stracke</td>
<td>Composition of the depleted mantle</td>
</tr>
</tbody>
</table>

Supplementary materials include data source list.

~400 papers listed.

This list is not counted in any citation index.
Citations of Databases, and some data papers


**Figure 1.** OIB and MORB data are from the GEOROC and PetDB databases\(^4\), respectively …

List of papers with noble gas data used listed in Supplement,

but not the sources of the other chemical and isotope data.
Editors Breakout GERM 2006

- Citations:

- Problem: Databases are being cited more and more rather than the source publications

  - Need for citation of the data papers as well as the database.

  - Search needs to yield a reference list that can be attached to the publication for references.

  - Citation most effective if included ISI database
ISI Policy:

ISI refuses to count any citations in supplemental materials.
Summary

Increased use of geochemical databases means original work often not cited.

Solutions?

Pressure ISI to change procedures?
New ways to apportion credit?