Published bi-annually by the
MARGINS Office
Lamont-Doherty Earth Observatory
of Columbia University
61 Route 9W, P.O. Box 1000
Palisades, NY 10964-8000
USA

MARGINS

Newsletter No. 9, Fall 2002

Deposition from very high sediment yield New Zealand rivers is captured in upper margin basins

Alan R. Orpin[1], Lionel Carter[1], Steve A. Kuehl[2], Noel A. Trustrum[3], Keith B. Lewis[1], Clark R. Alexander[4], and Basil Gomez[5]

1. NIWA, PO Box 14-901, Kilbirnie, New Zealand; E-mail: a.orpin@niwa.cri.nz
2. Virginia Institute of Marine Science, School of Marine Science, College of William and Mary, Gloucester Pt., VA 2306, USA
3. Landcare Research Ltd., Private Bag 11052, Palmerston North, New Zealand
4. Skidaway Institute of Oceanography, 10 Ocean Science Circle, Savannah, GA 31411, USA
5. Geomorphology Laboratory, Indiana State University, Terre Haute IN 47809, USA

By any measure, the rivers draining the Ruakumara Peninsula of New Zealand are incredibly muddy (Figure 1). Despite their obscurity and relatively small size, the Waipaoa, Uawa and the Waiapu Rivers yield a massive 6, 800, 13,890, and 20,520 t/km^2/yr of suspended sediment respectively, among the highest yields anywhere on Earth. The importance of mountainous, small-catchment, high-yield rivers to the global sediment budget has been emphasised by Milliman and Syvitski (1992). With the exception of the Niger River, no other upland, low-land or coastal plain river has an annual sediment load >20 Mt, even though more than 25 such rivers have drainage basin areas over 60 times greater than that of the Waiapu. The Niger's drainage basin is some 700 times greater in area than the Waiapu. Collectively, the annual sediment output of the Waipaoa, Uawa and the Waiapu Rivers represents about 0.3% (~60 Mt) of total global input to the oceans. They drain a mountainous landscape largely of unconsolidated Tertiary mudstone, where erosion has accelerated 5-10 fold since European deforestation.

Together with its high sediment flux, the Ruakumara margin has other features that make it a compelling setting for research: (1) the three rivers have century-long flow records; (2) the dramatic and well documented acceleration of erosion resulting from deforestation since Polynesian and European settlement (c.800 and c.150 years ago respectively); (3) convergent margin tectonism that produces rapidly subsiding mid-shelf basins to trap most terrigenous sediment input; (4) three large re-entrants caused by slope failure that incise the margin and narrow the shelf significantly, the bathymetry of which have now been surveyed using high-resolution multibeam; and (5) a marine stratigraphic record punctuated by numerous dated tephras. Accordingly, this margin serves as a reference for high terrigenous sediment flux and shelf entrapment, for anthropogenic effects on a virgin system, and for sediment-tectonic interactions on a steep and unstable continental slope.

These features were critical for the Waipaoa's inclusion into the MARGINS Source-to-Sink program that proposes a holistic approach to the study or continental margins; see

http://www.ldeo.columbia.edu/margins

In This Issue:

Data Management in the Earth Sciences

This issue of the MARGINS Newsletter is devoted to the important topic data management in the earth sciences, and contains a thematic set of articles:

- *New NSF-OCE Data Policy*
- *Ridge 2000 Integrated Study Site Data policy*
- *WDC-MARE/PANGAEA: a convenient RDBMS partner with MARGINS?*
- *MG&G Database Efforts at Lamont-Doherty Earth Observatory*
- *Seismic Reflection Data Access at University of Texas Institute for Geophysics*
- *A GERM Perspective on MARGINS Databases*
- *Towards a Data Management structure for MARGINS: Examples from Boomerang 8 and the Virtual Research Vessel*

Table of contents for the data management articles is on **page 15**

Researchers from NZ and the U.S. are planning collaborative, interdisciplinary studies of the Waipaoa river system with a particular focus on the propagation of marked environmental shifts through the sedimentary system and their ultimate preservation in the stratigraphic record of the Poverty Bay shelf and slope.

Sediment architecture on a subsiding shelf

The Raukumara margin is composed of allochthonous Palaeogene rock with a Neogene sediment cover, and is tectonically-active because of oblique subduction. Convergence has imbricate-thrust and folded Neogene slope sediments as a deforming backstop, with only a narrow accretionary prism locally forming at the toe of the slope. Seaward of the three rivers, post-glacial deposition occurs predominantly in actively subsiding, mid-shelf basins with maximum sediment thicknesses exceeding 100 m off the Waiapu. Each basin is boarded at its seaward edge by a structural high formed by an anticline of Neogene mudstone that is emergent near the shelf-edge (Figure 2). Early interpretations of Quaternary seismic stratigraphic sequences in a mid shelf basin in Hawkes Bay, south of East Cape, demonstrated the interplay of eustatic sea level, local and rates of sedimentation, tectonic uplift and subsidence. This framework has been expanded and many of the regionally extensive reflectors are now widely characterised and dated. On the shelf, thin continuous reflectors dip at a low angle seawards but are upturned against the flanks of the growing anticlines at the shelf-edge. Hence, the anticlines provide an effective barrier to seaward progradation of Quaternary sequences offering an opportunity to date the seaward feather-edge of each marine sequence. The postglacial shelf architecture of the East Coast is characterised by two major reflectors that can be used to adequately define the sediment prism, namely the last post-glacial transgressive erosion surface and a conformable strong reflector in the top 15 m of the prism. The latter is tentatively considered to be early Holocene in age, and is commonly characterised by gas masking.

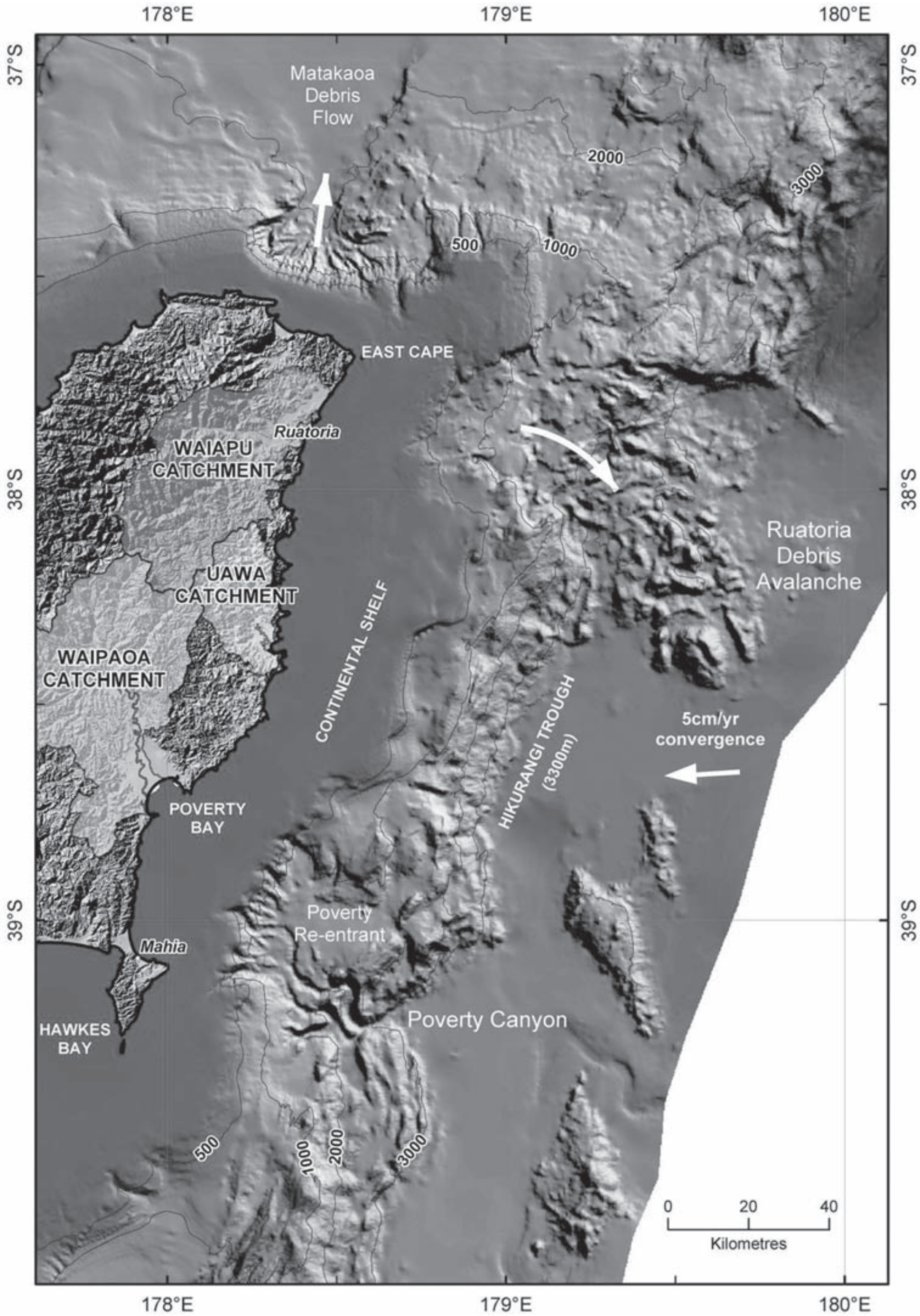

Figure 1. Topography and bathymetry of the Waiapu, Uawa, and Waipaoa Rivers, eastern North Island of New Zealand and the adjacent shelf areas.

Comparison with the recently studied Eel River margin, another tectonically active environment with a flood-prone river draining a mountainous catchment, emphasises the considerably higher specific sediment yield of the Waipaoa and the Waiapu Rivers (4 and 13 times greater respectively) and the higher sedimentation rate on the Eel margin slope, the result of only ~20% of the sediment remaining on the shelf (Sommerfield and Nittrouer, 1999). Sediment accumulation rates off the Waipaoa determined using ^{210}Pb geochronology indicate a rate of ~1 cm/yr on the outer shelf, with this rate being halved for the mid-shelf basin.

Hence, the modern sediment accumulation is inconsistent with the post-glacial sediment thicknesses, which show the largest volume accumulating on the mid-shelf. This discrepancy may reflect a modern reduction in middle shelf subsidence or a higher frequency of hyperpycnal flows from the Waipaoa River, with their ability to transport sediment seawards. In sharp contrast to the Eel River margin, accumulation rates on the Poverty Bay slope are an order of magnitude lower than the shelf at ~0.1 cm/yr.

Offshore from the mid-shelf depocentres are the large Poverty, Ruatoria, and Matakaoa margin re-entrants; the scars of massive debris flows and avalanches up to 3000 km^3 in volume that incise into the imbricate-thrust and uplifted margin (e.g. Collot et al., 2001). These re-entrants are also close to the mouths of the Waipaoa and Waiapu Rivers, providing preferential sediment pathways to the deep ocean via a complex network of small head-scarp gullies, collapse structures, and canyons. However, despite these conduits, provisional investigations suggest that limited sediment escapes capture on the upper margin. Cores and geophysical data suggest the majority of terrigenous sediment is trapped in shelf and upper slope basins, perhaps even during sea level lowstands. Such entrapment off the Waipaoa system has essentially starved the nearby Poverty Canyon system that feeds into the 3,300 m-deep Hikurangi Trough. Instead of leading onto a submarine fan, the canyon is terminated in a sediment starved depression with exposures of stiff Pleistocene mud.

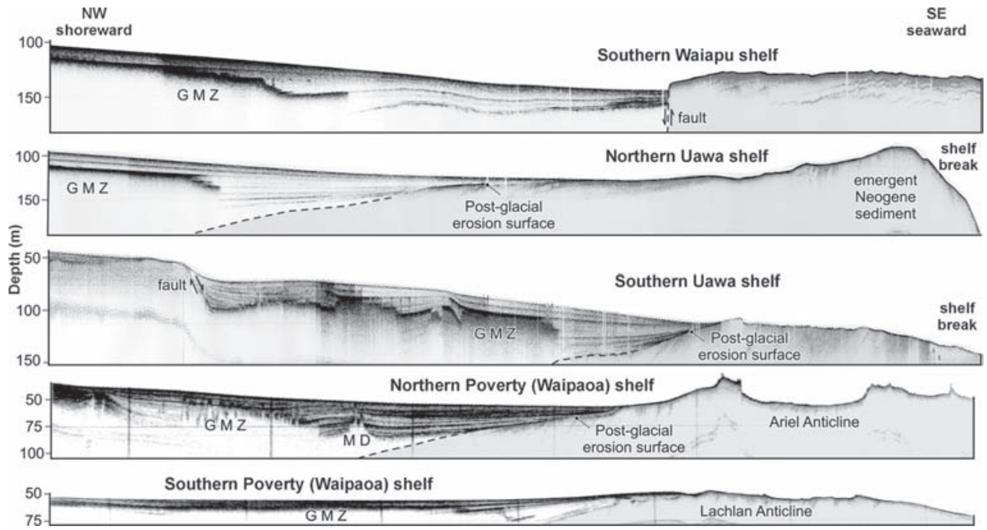

Figure 2. Seismic sections across the mid-shelf at the three rivers, showing mudstone structural high on the seaward side. MD = mud diapir, GMZ = gaseous mud zone

A history of dramatic land use changes

Polynesian settlers (Maori) started to clear the landscape of thick temperate rain-forest 800-500 ^{14}C yr BP, and forest clearing accelerated with European colonisation in the mid-eighteenth century. By 1880 most of the hinterland had been cleared, and by 1920 all but a few percent of the land had been converted to pasture. A particularly intense phase of erosion was initiated in the upper reaches of Waipaoa catchment around the turn of the 19th century following deforestation and the river has been aggrading in response to the increased sediment yield since (Gomez et al., 1999).

The even muddier Waiapu River shares many similarities with the Waipaoa including catchment size, rock lithology, climate, and deforestation history. Much of the hinterland of the Waiapu catchment had reverted to scrub by the 1950's. Following deforestation around 1920, large gully complexes developed and riparian landsliding increased. Today, the specific sediment yield for the Waiapu is 2.5 times higher than that of the Waipaoa, which is a reflection of the Waiapu's greater susceptibility to deep-gully erosion, a process that produces more than 50% of its total sediment yield. The Waiapu catchment has twice the number and area of active gullies compared to the Waipaoa, due partly to 60% greater rainfall, more frequent storms, and greater shearing and fracture of older source rocks (Page et al., 2001a and b).

Consistent with the behaviour of meso-scale river basins, which are too small to modulate rainfalls, high-intensity storms play an important role in sediment transfer to the coast. Storms trigger shallow landsliding on Tertiary rocks, whereas the crushed and fractured Cretaceous greywacke and argillite of the headwaters respond with extensive gully erosion. Early attempts to control gully erosion using check dams were ineffective. However, post-1960 reforestation of the most severely eroded headwaters has helped stabilize many active gullies. The resultant reduction in gully erosion has been accompanied by an increase in channel down cutting and a decrease in sediment load. Today, around 20% of the Waipaoa and Waiapu catchments are reforested with exotic pine.

Timing terrestrial events offshore

Cores from the shelf and upper slope contain a near-continuous stratigraphic record of post-glacial sedimentation. Two major events that are particularly well represented are the Taupo volcano eruption 1,718 ky BP and European deforestation. Pollen records show that immediately following the Taupo eruption there was a 13% increase in sedimentation that persisted for several centuries, and is related to both a patchy disturbance of podocarp rain forest and fluvial redistribution of tephra. These pollen data also indicate a possible temporal offset of decades up to a century between terrestrial vegetative landscape changes and the response in the marine record. Data from Foster and Carter (1997) and a 17 m Calypso core (MD2122) from the mid-shelf depocentre off the Waipaoa River suggested that European deforestation caused a four-fold increase in sediment accumulation.

Cyclone Bola in 1988, an event with a 100 year return period, caused record flooding throughout the Raukumara Peninsula. In the Waipaoa River, the suspended sediment load for the 6 day event exceeded 32 Mt, more than double the annual mean suspended load (Hicks et al., 2000). River sediment was discharged to the coast as a seabed-hugging hyperpycnal flow, which generated an organism-smothering fluid mud layer. Reef communities on the inner shelf were inundated and at the mouth of the Waipaoa River, Gisborne Harbour entrance experienced rapid infilling more than 20 times the average rate. However, despite the severity of the flood and mass of sediment delivered to the shelf, coastal accretion reverted back to a delicate equilibrium state within two years. Anecdotal evidence suggested that the fluid-mud layer was up to 2 m thick in places, but was resuspended and dispersed, or mixed into the sea floor, within weeks after the flood. In the absence of field measurements, the frequency of these hyperpycnal flows remains speculative, but continuous gauging data from the lower reaches of the Waipaoa suggest suspended-sediment concentrations are high enough (>20 gL^{-1}) to frequently trigger hyperpycnal flows at the river mouth several times annually.

To date, discrete storm events such as Cyclone Bola have not been fingerprinted conclusively in marine cores although lacustrine records, such as those recovered from Lake Tutira in the northern Hawkes Bay hinterland, show a high resolution history of 365 storms and climate cycles for the last 2,250 years (Eden and Page, 1998). On a decadal time scale, the most recent reduction in sediment derived from reforested hillslopes has yet to be identified downstream at the basin outlet or offshore. The implication is that there is a lag in sediment delivery due to significant volumes of erodible sediment being stored in the upper and middle reach channels. A major component of the ongoing and future (i.e. MARGINS) research programs is to trace these process-response perturbations and assess how they are manifest in the marine record and are propagated through the terrestrial-marine system.

Sediment pulses associated with floods or volcanic eruptions can serve as a proxy to fingerprint the sedimentary system, giving vital clues as to the response times for severe catchment degradation and recovery, changes in sediment composition and texture, and the effect of floods. In contrast to many settings, supply and accommodation space are not limiting factors along the Raukumara Peninsula margin, thus making it a prime site for analysis of Holocene climate change through the reconstruction of flood and other meteorologically forced events.

Acknowledgements

This work is funded by the New Zealand Foundation for Research Science and Technology grants C0X0013, C01X0037, C01X0038, and a NZST Postdoctoral Fellowship NIWX0003.

References

Collot, J.-Y., Lewis, K., Lamarche, G., and Lallemand, S. (2001). The giant Ruatoria debris avalanche on the northern Hikurangi margin, New Zealand: result of oblique seamount subduction. Journal of Geophysical Research 106, 19271-19297.

Eden, D.N. and Page, M.J. (1998). Palaeoclimatic Implications of a storm erosion record from late Holocene lake sediments, North Island, New Zealand. Palaeo 139, 37-58.

Foster, G. and Carter, L. (1997). Mud sedimentation on the continental shelf at an accretionary margin – Poverty Bay, New Zealand. New Zealand Journal of Geology and Geophysics, 40:157-173.

Gomez, B., Eden, D.N., Hicks, D.M., Trustrum, N.A., Peacock, D.H. and Wilmshurst, J.M. (1999). Contribution of floodplain sequestration to the sediment budget of the Waipaoa River, New Zealand. In, Floodplains: Interdisciplinary Approaches (S.B. Mariott and J. Alexander, Eds). Geological Society of London Special Publication 163, 69-88.

Hicks, D.M., Gomez, B. and Trustrum, N.A. (2000). Erosion thresholds and suspended sediment yields, Waipaoa River Basin, New Zealand. Water Resources Research 36, 1129-1143.

Milliman, J.D. and Syvitski, P.M. (1992). Geomorphic/tectonic control of sediment discharge to the ocean: the importance of small mountainous rivers. Journal of Geology 100, 525-544.

Page, M., Harmsworth, G., Trustrum, N., Kasai, M. and Marutani, T. (2001a). Waiapu River (North Island, New Zealand). In, Source-to-Sink Sedimentary Cascades in Pacific Rim Geo-Systems (T. Marutani, G. Brierley, N. Trustrum, N. and M. Page Eds) p.102-111.

Page, M., Trustrum, N., Brackely, H., Gomez, B., Kasai, M., and Marutani, T. (2001b). Waipaoa River (North Island, New Zealand). In, Source-to-Sink Sedimentary Cascades in Pacific Rim Geo-Systems (T. Marutani, G. Brierley, N. Trustrum, N. and M. Page Eds) p.86-100.

Sommerfield, C.K. and Nittrouer, C.A. (1999). Modern accumulation rates and a sediment budget for the Eel shelf: a flood-dominated depositional environment. Marine Geology 154, 227-241.

From the Chairman's Desk: Fall 2002

Garry D. Karner

Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, New York 10964, USA; E-mail: garry@ldeo.columbia.edu

The Spring-Fall period of 2002 has not been particularly gracious to the MARGINS Program, with yet another upheaval with the Source-to-Sink initiative and a Californian court injunction preventing the R/V Maurice Ewing from continuing its seismic and underway geophysical acquisition in the Gulf of California, thus significantly compromising the objectives of the first major program of the Rupturing Continental Lithosphere initiative. However, some successes have been scored, such as the excellent Izu-Bonin-Mariana (Hawaii, September) and Subduction modeling (Michigan, October) workshops and the establishing of working agreements between Egyptian universities and institutes with U.S. researchers. Funding has also been approved for the second MARGINS Theoretical and Experimental Institute for March, 2003 and a major international workshop in New Zealand, May 2003, which will be used to summarize the current research efforts and results associated with the Waipaoa source-to-sink system and the preparation of collaborative proposals for the next MARGINS proposal deadline.

MARINE MAMMALS AND THE MARGINS PROGRAM: A NEW "PANDORA'S BOX" FOR EARTH SCIENTIFIC RESEARCH

On the morning of September 24th, two Cuvier's beaked whales beached themselves in the Gulf of California. The R/V Maurice Ewing, involved in multichannel seismic and underway geophysics data acquisition as part of the MARGINS Rupturing Continental Lithosphere Gulf of California project, was 30-40 miles away at the time of the beachings. When the ship operator (Columbia University) learned about this event, the ship's operations were shut down for several days while the operator (in consultation with NSF) investigated the situation and added even more rigorous measures to the vessel's existing procedures to insure that marine mammals were not impacted by seismic operations. Such measures included aerial overflights, no night time operations, reduced airgun volume, additional skilled marine mammal observers and restrictions concerning operations in shallow water. How close the vessel had been to the two whales was unclear. The date of the whales' deaths also was not known. Any causal link between the Ewing's operations and the death of the whales was unclear and undemonstrated and considered by many to be highly unlikely.

Nevertheless, the Center for Biological Diversity, an environmental group based in Tucson, filed a lawsuit seeking to stop the research. The group charged that the Ewing was probably harming sea creatures, and was violating the National Environmental Policy Act and the federal Marine Mammal Protection Act. On September 28th, the U.S. District Court for the Northern District of California issued a temporary restraining order, bringing the cruise to an abrupt halt. The situation remains unresolved.

Because the latest phase of the project has now ended, the lawsuit is unlikely to resolve the issues surrounding the research, as the court is likely to declare the suit moot. However, the legal issues will surely arise again when further phases of Ewing research take place, in late 2003 or early 2004. NSF, NOAA and the ship operator are in discussions in order to develop working procedures for conducting marine geological and geophysical operations in the world's oceans. These inter-agency discussions are absolutely crucial to the future of MARGINS and thus our science — the MARGINS Office will attempt to keep the scientific community informed of these deliberations and their outcomes.

PROCEDURES FOR CONDUCTING RESEARCH IN FOREIGN COUNTRIES

At this time in the MARGINS Program, it is timely to review the procedures required to set up collaborative programs between U.S. researchers and their foreign counterparts for the various countries within the MARGINS program.

Collaborations between individual researchers

U.S. researchers writing proposals to work in Costa Rica, Mexico Nicaragua, New Guinea or New Zealand need to recognize and contact individually their equivalent in these countries and include them as associate investigators in the NSF proposal. A letter from the foreign university department or institute outlining the form, conditions and expectations of the collaboration is all that is needed (at the proposal stage). All of these countries require visas and special permissions to work on land and in marine regions. For example, most U.S. researchers who conduct fieldwork in Mexico are doing so under the auspices of a tourist visa when the work is carried out on mainland Mexico or on the peninsula of Baja California. However, they really should have an appropriate immigration permit, granted by the Mexican immigration authorities. The following guidelines, while specifically for Mexico, have much in common with other countries involved in MARGINS research. To give a flavor of what is involved, I am including the following Mexican information for obtaining a visa and other work documents:

1. Obtain a letter of invitation from a Mexican university or research institute where their research is in com-

mon with the field of interest of the foreign researcher. In this letter it is necessary to mention the kind of relationship between the foreign and national collaborator, and also give a short description of the project as planned, as well as the benefit of the project to Mexico. It is important to mention that while the foreign researchers are in Mexico they will not receive any salary from any Mexican organization.

2. With the letter of invitation and a valid passport, the next step is to go to the nearest Mexican embassy or consulate to request immigration form to be a visiting professor. The consulate should grant this request, although you need to be aware that there are some nationalities that are restricted under Mexican immigration law. In those cases, the person has to ask for a special permit that is granted only by the Secretaria de Gobernación (Interior Department) in Mexico City, which usually takes several weeks to obtain if it is requested directly at their office in Mexico City, aided by a Mexican research institute or university. However, it can take many months if the process is done through a Mexican Embassy or consulate without any help from a Mexican institution.

3. There are areas within the country that are protected, such as national parks, biosphere reserves and the islands, where anyone (national or foreigner) who wishes to do any kind of field work has to have a permit by at least two federal secretaries, such as the Secretaria del Medio Ambiente, Recursos Naturales y Pesca (Secretary of Environment, Natural Resources and Fishery) and the Secretaria de Gobernacion (Dept of Interior). When work is to be done on an island, the Secretaria de Marina (Mexican Navy) also has to be notified.

Each researcher needs to solve these various visa and work permit requirements in consultation with their foreign counterpart. If there is already a growing experience with various countries, the MARGINS Office would be grateful if new-found knowledge could be shared with the office, so that the information can be summarized in the appropriate focus site web pages.

Collaborations with Egypt universities

In order for U.S. researchers to collaborate with Egyptian colleagues, they need to first generate a Memorandum of Understanding (MOU) between the respective universities and/or institutes - from the MARGINS Office experience, this seems to be the most effective procedure. The MOU outlines in a formal and legal way the general form of the collaboration and the conditions and expectations of the collaboration. This document must also define the various data sets that will be acquired and will form the basis of a proposal to the Egyptian Ministry of Science seeking permission or protocols for the acquisition and use of these data sets.

Collaboration between U.S. researchers and the Saudi Geological Survey

The Saudi Geological Survey has opted for a different model to encourage collaboration between U.S. researchers and the SGS. A general MOU was signed late last year between the MARGINS Office and the SGS, acknowledging the common research interests between the SGS and the objectives of the MARGINS Rupturing Continental Lithosphere initiative. The MOU acknowledges that the SGS is the primary geologic survey agency in the Kingdom and the principal repository for Earth Science data. The MOU also outlines a number of agreements that will help facilitate collaborative research between U.S. and Saudi scientists:

- It was agreed that the Saudi Geological Survey (SGS) will act as the Saudi point-of-contact between the MARGINS Office and U.S. researchers in helping to facilitate invitations for visas, helping with field logistics and providing assistance for obtaining permits and clearances for onshore and offshore surveys, respectively.
- It was agreed that all data acquired under the auspices of the MARGINS program will be shared by all collaborators and can be freely published in appropriate formats in internationally recognized scientific journals. After two years, these data become public information (consistent with MARGINS data policy).
- It was agreed that MARGINS projects should attempt to incorporate high-priority SGS projects where possible.
- It was agreed that all MARGINS projects and investigations in Saudi Arabia or Saudi waters should be collaborative projects involving both U.S. and Saudi researchers.
- It is understood that visits of Saudi researchers and students to U.S. institutions and of U.S. scientists and students to Saudi institutions is an important element in these collaborative projects. The MARGINS Office and the SGS will both work to encourage and facilitate such exchanges.

It is also important to note that the signed Memorandum of Understanding on future collaborative research does not obligate the MARGINS Office or Program for financial support of anticipated collaborative research, expressing only the intent that proposals seeking such support from NSF and Saudi funding agencies will be generated for collaborative research by U.S. and Saudi researchers.

THE MARGINS STEERING COMMITTEE

Two people rotated off the committee during the Spring, 2002 semester. The committee would like to take this opportunity to thank Simon Peacock and Doug Wiens, for their unselfish contribution of time and effort to the steering committee and the Earth Scientific community. Simon and Doug both provided sound judgment and advice during the deliberation of a range of committee and community issues, most importantly on the fellowship and data policy documents, respectively. The committee thanks Simon also for his willingness to represent his community on being a co-author on a number of SubFac workshops.

In turn, the committee would like to welcome the following new members onto the steering committee: Geoff Abers and Mark Reagan. Geoff brings to the committee valuable expertise concerning passive seismology and tomography and will be an important voice now that Doug has rotated off the committee. His interests in the IBM and Costa Rica/Nicaragua focus sites will be important in representing the SEIZE and SubFac communities on the committee. Mark has expertise and an intimate knowledge of arc geochemistry and geochronology issues as applied to Costa Rica, Nicaragua and the Marianas. He will be primarily representing the SubFac community.

Last but certainly not least, it gives me great pleasure to announce that Julie Morris will be the next Chair of the MARGINS Office, effective 1 October, 2003. During July-September, 2003, the MARGINS will undergo a slow transformation from Lamont-Doherty (New York) to Washington University (Saint Louis) as activities, responsibilities and web databases are transferred to the new office.

NSF-IFREE-MARGINS Workshop on the Izu-Bonin-Mariana Subduction Factory (IBM 2002)

James B.Gill[1], Simon Klemperer[2], Robert Stern[3], Yoshi Tamura[4], Douglas Wiens[5]

1. Earth Sciences Department, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA; E-mail: jgill@es.ucsc.edu
2. Geophysics, and Geological & Environmental Sciences, Stanford University, 397 Panama Mall, Mitchell Building 353, Stanford, CA 94305-2215, USA
3. Center for Lithospheric Studies, University of Texas at Dallas, PO Box 830688, 2601 N. Floyd Road, Richardson, TX 75080, USA
4. Department of Earth Sciences, Kanazawa University, Kanazawa 920-1192, Japan
5. Department of Earth and Planetary Sciences, Washington Uiversity, 1 Brookings Drive, St. Louis, MI 63130-4899, USA

A. Overview

A workshop on the Izu-Bonin-Mariana (IBM) subduction system was held in Honolulu HI during September 8-12, 2002, under the auspices of the MARGINS Program of NSF. It was co-sponsored by the Japanese Institute for Frontier Research on Earth Evolution (IFREE). Both the USA and Japan have selected the IBM for focused research during the next five to ten years, creating opportunity for joint research activities.

Convenors in alphabetical order were J. Gill (Santa Cruz), S. Klemperer (Stanford), R. Stern (Dallas), Y. Tamura (IFREE), and D. Wiens (St. Louis). About 100 scientists attended. About 2/3 were from the USA and most of the rest were from Japan. Many had not worked extensively in the region before, and about 25% were graduate students and postdocs from both countries. There were about twenty invited talks, listed below, about equal time for discussion, and about 50 poster presentations. Abstracts and key visuals from the invited talks are available at:

http://www.ldeo.columbia.edu/margins/SubFac/IBM/IBM02.html

The IBM arc has been selected as the "oceanic cold-subduction" end-member example in which Subduction Factory topics can be addressed effectively with least influence of the upper plate. It is one of two integrated study sites for such projects; Central America is the other. The workshop featured recent research results related to IBM, some of which were sponsored by the MARGINS Subduction Factory initiative. Similarities and differences between the two study sites also were presented, as were comparisons with other oceanic arcs.

The meeting started with overviews of the geochemical and geophysical context of general subduction by C. Hawkesworth (Bristol) and M. Gurnis (CIT), and of the IBM arc in particular by R. Stern (UTD) and B. Taylor (UH), respectively. M. Arima (Yokohama) summarized the IBM arc components exposed in the Tanzawa arc-arc collision complex. S. Peacock (ASU) discussed the thermal and flow structure of IBM, and G. Abers (BU) described general slab seismicity and seismological constraints on the dehydration and phase transformations in the downgoing crust. D. Wiens (WUSL) and M. Fouch (ASU) described the seismic tomography and anisotropy of oceanic arcs in general as background to work in the Marianas in 2003. S. Klemperer (Stanford) and K. Suyehiro (JAMSTEC) summarized the crustal structure of the Mariana and Izu sectors, respectively.

The second day focused on inputs and outputs to the IBM arc. T. Plank (BU) presented results from ODP Leg 195 for slab inputs, and R. Hickey-Vargas (FIU) summarized mantle inputs from the perspective of back arc basin basalts. P. Fryer (UH), T. Elliott (Bristol), and O. Ishizuka (GSJ) summarized outputs from the forearc, volcanic front, and backarc, respectively. Y. Tamura focused on the abundant felsic volcanic and plutonic outputs of Izu, and J. Ishibashi (Kyushu) summarized IBM's hydrothermal fluids and deposits, and their geobiology. M. Reagan (Iowa) summarized the history of IBM's magmatic outputs, and K. Fujioka (JAMSTEC) provided a tectonic overview and estimate of crustal production rates.

The third day included reviews by M. Hirschmann (Minnesota) of decompression and flux melting in arcs, and by B. Bourdon (IPG-Paris) of the timescale of such processes from the perspective of U-series disequilibria. P. Kelemen (WHOI) discussed processes generating continental crust in oceanic arcs, especially the Aleutians. S. Schwartz (UCSC) discussed the shallow seismogenic zone and some of the unique observations of that zone in the Izu and Mariana arcs, G. Hirth (WHOI) discussed the effects of water on the material properties of the mantle wedge, and Y. Tatsumi (JAMSTEC) gave an overview of the multiple roles of subduction in crustal and mantle evolution. Summaries were also given of NSF, IFREE, and GEOMAR plans for work on subduction zone processes in IBM and Central America, and the related RIDGE2000 integrated study site in the Lau Basin. Plans for IODP drilling and submersible programs in IBM were presented. The day closed with brief descriptions of about a dozen funded field projects in the IBM system.

B. Progress towards Realizing the Objectives of the Subduction Factory Science Plan

Results were presented from, or in preparation for, all six of the MARGINS Subduction Factory projects that have been funded in the Mariana portion of the IBM arc since 1999. They included >5000 km of multi-channel seismic profiling (Taylor et al.), a 50-OBS wide-angle seismic experiment for crustal structure (Klemperer et al.), passive seismic tomography (Wiens et al.), melt inclusions in phenocrysts (Plank et al.), age dating and geochemical evolution of islands (Reagan et al.), and the southern seamount province (Stern et al.). In addition, results were presented from ODP Leg 195 which drilled sediments and basaltic basement in the incoming Pacific Plate off both the Izu and Mariana arcs.

The workshop demonstrated that progress is being made in the IBM focus site towards providing good answers to the questions that motivate the Subduction Factory Initiative: 1) How do forcing functions such as convergence rate and upper plate thickness regulate production of magma and fluid from the Subduction Factory? 2) How does the volatile cycle (H_2O and CO_2) impact chemical, physical and biological processes from trench to deep mantle? 3) What is the mass balance of chemical species and material across the Subduction Factory, and how does this balance affect continental growth and evolution? In addition, one of the ancillary questions asked in the Science Plan: “How, why and where are new subduction zones started?” seems to be best addressed at this site, and good progress is being made to understand this process as well. Examples of the progress are provided below.

1. How do forcing functions such as convergence rate and upper plate thickness regulate production of magma and fluid from the Subduction Factory?

A major difficulty in understanding subduction processes has been inadequate numerical models for temperature and mantle flow in subduction zones. Three new thermal models (Conder et al., Peacock and van Keken, and Kelemen et al.) were presented at this meeting that incorporate temperature-dependent viscosity. These models showed higher slab surface temperatures than previous models, and in addition suggested a possible mechanism for a component of decompression melting beneath the arc. Workshop participants also learned about the first-order subdivision of the IBM forearc into a deforming southern (Mariana) part and a relatively undeformed northern (Izu) part. It seems clear that the abundance of serpentinite diapirs and forearc vents are controlled by this deformation. The lithospheric structure of the upper plate of the Izu segment has been determined by Japanese scientists and a parallel study in the Marianas funded by NSF has begun. A long active source profile across the Mariana arc with 100 ocean bottom seismographs will be completed by Japanese scientists in early 2003. Forcing functions related to mantle flow and possible sequential melting are being investigated by Japanese and US investigators by studies linking magmatic outputs along flow lines from the backarc basin spreading ridge along cross-chains and into the magmatic front. Waveform inversion results for the structure of the Mariana Trough backarc summarized by D. Wiens suggest higher upper mantle seismic velocities and possibly lower temperatures than in the Lau Basin backarc. No high resolution tomography has been completed in IBM yet but this is one of the goals of the funded passive OBS deployment to commence in 2003. We still do not understand what is causing the first-order variation in magmatic compositions along the arc, from moderately enriched (medium-K calc-alkaline) in the south to extremely enriched (shoshonitic) in the center to ultra-depleted (low-K tholeiitic) in the north. Part of the answer comes from the different types of sediments that are being subducted, as T. Plank showed at the workshop, but this does not explain the observation that Izu arc lavas are higher degree melts than those of the Mariana arc. An important model for explaining why arc melts are so depleted is that they are products of sequential melting, first beneath the back arc, then beneath the volcanic front. This model is difficult to reconcile with the fact that lavas from the Mariana Arc, which are associated with an actively spreading back-arc basin, are

Figure 1. Scenes from a workshop: Jim Gill opens the workshop (top left). Discussing inputs and outputs from the Subduction Factory in the Imperial Ballroom (top right). The participants got a warm, although windy reception by the MARGINS Office and the conveners (bottom left). The bottom right picture shows the workshop participants before the palms of Waikiki Beach.

less depleted than Izu Arc lavas, which are not. The slab component also seems to differ along strike, with "fluid" and "melt" signatures being combined at the Marianas volcanic front but partitioned between volcanic front and reararc, respectively, in Izu. Workshop participants also saw new GPS results (T. Kato et al.) for the Mariana Arc which radically alter our understanding of how the Mariana Trough is opening, and it is clear from this that GPS studies of plate motion are essential.

2. How does the volatile cycle (H_2O and CO_2) impact chemical, physical and biological processes from trench to deep mantle?

Recent improvements in microanalytical techniques (ion probe, laser ablation ICP-MS, FTIR) have resulted in tremendous advances in understanding element fluxes, especially water, through arc systems, and the IBM system has been the site of many of these advances. Good data sets for water through the Mariana magmatic system are being assembled, and comparable data sets for glass inclusions in Izu ejecta are needed. Initial results suggest higher water contents in Mariana than Izu magmas of similar level of differentiation. The flux of carbon dioxide through the IBM arc magmatic system is controversial, with no good data yet because hydrous magmas lose this gas even at relatively high pressure. Obtaining robust estimates of CO_2 flux through convergent margins is a global challenge. Studies of subaerial fumaroles and submarine hydrothermal and forearc vents are needed as part of experiments designed to monitor the fluxes. Estimating sulphur dioxide fluxes is tractable but remains to be widely accomplished. Monitoring SO_2 fluxes can be done to advantage using satellite remote sensing, so that involving NASA and NASDA (Japan space agency) in this effort is highly desirable. Measuring the flux of other volatiles (halogens, N, methane, rare gases) and related isotopic compositions is also needed, and these measurements would offer important insights into CO_2 flux. Workshop participants saw good examples of how these measurements were progressing in Central America (T. Fisher, D. Hilton, etc.), and this approach could be adapted to IBM. IBM has an additional flux from hydrothermal systems associated with backarc basin spreading system as well as submarine calderas, and techniques developed as part of the RIDGE program could be readily adapted. Japanese scientists are very active in measuring volatile fluxes both at sea and on land and opportunities for interaction exist. Work on how these fluxes affect the biosphere is just getting underway and US investigators may learn a lot from the Japanese 'Archean Park' project just getting underway in southern IBM. This project is led by T. Urabe (University of Tokyo) with co-investigators from a number of Japanese institutions including JAMSTEC; US investigators should be encouraged to collaborate.

3. What is the mass balance of chemical species and material across the Subduction Factory, and how does this balance affect continental growth and evolution?

Results from ODP Leg 185 were presented by T. Plank, who showed that some of the variation in incompatible elements along the IBM arc system can be related simply to variations in the composition of subducted sediments. IBM is a very good arc system to identify input controls, because the sediments being subducted in the north and the south are distinct. Whether or not this model can explain isotopic variations along the arc remains to be resolved. The magnitude of magmatic fluxes, and whether they vary systematically along the arc, remain important unresolved questions. Because crust formation in arcs is largely vertical, this flux cannot be related to a measurable quantity such as spreading rate. Consequently, it is difficult to measure the magmatic flux directly. Techniques for estimating magmatic flux indirectly (i.e., via SO_4 or some other volatile flux) await development. The thickness of the crust, which integrates the arc magmatic flux over the life of the arc, can be measured directly, and good progress is being made in IBM. Japanese scientists have already measured crustal seismic velocities across the Izu arc and interpreted this for crustal structure. The field experiment needed to generate a comparable profile along the Mariana arc has been conducted by S. Klemperer et al. and preliminary results were presented at this workshop. Further complementary work is being planned by JAMSTEC scientists. There was extensive discussion about the significance of the 6.2 km/sec P-wave velocity layer observed in the Izu cross-section, and the likelihood that tonalites exposed in the Tanzawa mountains in the Izu collision zone represent exposures of that layer. This felsic middle crust is similar in some aspects to the composition of continental crust but dissimilar in trace element concentrations and isotopes, being more depleted in both compatible and incompatible elements. The crustal structure and composition of the two intra-oceanic arc systems for which there are high quality crustal velocity profiles, Izu and the Aleutians, are so different as to require more studies to determine the significance of this variability. Workshop participants were intrigued by the likelihood that arc lower crust delaminates and falls back into the mantle, as has been suggested for the Sierra Nevada of California, but understanding this process remains an important challenge for studies of crustal growth.

4. How, why and where are new subduction zones started?

Good progress is being made here. The workshop was excited to see geodynamic models developed by C. Hall and M. Gurnis (CIT) which reproduces the most important features of the IBM subduction initiation process. Uncertainties about the paleogeography of the West Philippine Basin continue to complicate our understanding of how subduction began. R. Hickey-Vargas demonstrated that lavas from the West Philippine Basin are similar to those of the Indian Ocean. M. Reagan showed new geochronologic results for proto-arc igneous activity, along with compositional data for these lavas, but results remain too scattered to resolve genetic models.

Another important objective of the workshop was to encourage collaboration between US and Japanese scientists working on related problems. Several such joint opportunties were discussed. 1) IODP Drilling in IBM. Several specific ideas were addressed; see below. 2) Comparative field studies. Despite sharing a common history, IBM sectors differ in important ways. For example, slab dip steepens southward, there is backarc rifting in the north but spreading in the south, and “fluid” and “sediment” components are combined in arc magmas in the south but separated across-arc in the north, at least since 15 Ma ago. Along-strike comparisons offer powerful opportunities to evaluate differences in forcing functions, volatile behavior, and mass balances, but require joint planning and sometimes joint deployment of assets. Several joint field programs have already been planned and funded, including a joint passive land-OBS deployment aboard the Japanese ship Kaiyo to commence in the Marianas in 2003 (Wiens and Suyehiro/Shiobara). The US (Klemperer) and Japanese (Suyehiro/Kodaira) active source experiments in the Mariana have been coordinated and should yield complementary information. Further joint field programs were discussed. 3) Numerical model of subduction zones. A subduction modeling workshop, funded by MARGINS and held in Michigan in October, should rapidly increase progress in this area. The Earth Simulator at IFREE, currently the most powerful computer available to the scientific community, is available to run three-dimensional kinematic models of subduction at high spatial and temporal resolution. Participants felt that models are more limited conceptually than computationally at the moment, but joint efforts in this field could change this quickly.

C. Topical summaries.

The meeting ended with a full day of small group discussions of specific topics to identify what still needs to be done and how to do it, in light of the oral and poster presentations and related interaction of the preceding days. The topics were chosen by participants and are not mutually exclusive. Brief summaries of these discussions are on the Workshop website given above. The topics were: 1. Fluids and melts from the slab; 2. Imaging, modelling, and experimenting on mantle wedge processes; 3. Timescale of dehydration and melting processes; 4. Volatile fluxes and cycles; 5. Nature and distribution of primitive melts in the mantle wedge; 6. Subduction initiation; 7. Crustal evolution and intermediate/felsic magmas; 8. Critical geologic, geochemical, and geophysical data still required; 9. IODP planning; 10. Future off-shore projects; and 11. Future on-land projects.

Most small group discussions identified topics they considered most important for future research. Not surprisingly, it was concluded that most pivotal subduction zone processes can be addressed well in the IBM system. Its along-strike and across-strike variations offer limitless opportunity. This is complemented by results of earlier ODP drilling at 18°N and 32°N which provide an unparalleled temporal history of arc volcanism and paleogeographic history at the same locations that crustal velocity structures are being obtained. However, in order to realize Subduction Factory objectives, the following kinds of additional work are still needed.

1. Geological and geophysical field programs. On the more geological side, there is need for further sampling and study of forearc serpentinites and their fluids, and of crustal-level differentiation processes at arc volcanoes. The latter is made difficult by the remoteness of the islands, but several Izu volcanoes have well-established histories. There is also opportunity for sampling deep crustal sections and even mantle at the largest fault scarps. On the more geophysical side, there is need: to determine at least one more crustal velocity structure in the thinnest part of the arc; to deploy OBSs to determine Vs and Vp tomographic images of the mantle wedge to a resolution of ~25 km in the southern, middle, and northern segments above the shoaling slab; to mount GPS campaigns to determine mechanisms and rates of backarc rifting and spreading; to determine heat flow across the arc; and to deploy OBEMs to assess the distribution of melt and fluid in the mantle. An ARC-MELT sized program seems warranted and feasible, and is made more effective by the abundance of local deep earthquakes. In addition, there is need and opportunity to study at least one arc volcano from its roots, through its crustal level “magma chamber(s)”, to its orifice, similar to recent images of the axes of mid-ocean ridges.

2. Seismology programs. In addition to the above, emphasis also was given to the need to refine the spatial resolution of anisotropy, obtain more accurate hypocentral locations for earthquakes in the slab, and use those earthquakes to constrain the velocity structure of subducting crust and the presence or absence of fluid/melt at depths of 50-250 km beneath the arc. A better understanding of the lack of large seismogenic zone earthquakes in IBM and the possible role of different mechanisms, such as temperature, fluids, slab topography, and serpentinization, in limiting the size of thrust earthquakes is necessary. Progress can be made through comparison of accurate earthquake locations with higher resolution crustal structure.

3. Experimental programs. Because of the importance of slab dehydration for many subduction zone processes, many groups called for further experimental study of solid/fluid partition coefficients. In particular, there is need for study at controlled fO_2, CO_3, and halogen contents because the pioneering work to date has not converged enough to support quantitative modeling. In addition, more work is needed to resolve uncertainty about the water-saturated and damp peridotite solidii and near-solidus melt compositions, especially for depleted compositions. Similar experimental work is needed to relate seismic observables (e.g., V_s, V_p, Q, anisotropy) to physical properties of the upper mantle and lower crust. Better experimental and field constraints on the development of seismic anisotropy are particularly needed.

4. Numerical modeling programs. There was discussion of whether a "Community Numerical Model" of the IBM subduction zone(s) was a long term goal. Many thought so, but differed on the best way forward at this time Most concurred that such a model would be feasible in the future, but much work has to be done first. The Michigan subduction modeling workshop (see workshop report on page 13) may be a first step to developing cooperation between the different modeling groups and approaches. Access to the Earth Simulator should eventually allow higher resolution models to be developed. Geochemical and geological as well as geophysical parameters and observables should be included. Modeling of specific events, including subduction initiation and periodic backarc spreading are related but more specialized objectives.

5. Analytical programs. Need was especially recognized for intensive work on: melt inclusions in phenocrysts in rapidly quenched scoria to determine pre-eruption volatile concentrations; the positive correlations between the amount of slab components and the degree of melting; comparison of slab and mantle components along and across strike; integrated multi-nuclide U-series disequilibria to reconcile their time scale information about melt formation and differentiation; and rare gases across and along all arc sectors.

6. Geological and geochemical history programs. Better constrained paleogeographical reconstructions and radiometric ages are needed to enable improved understanding of subduction initiation. Further analytical study of volcaniclastic sediments (ashes and tubidites) is needed to understand why arc outputs differ between the Eocene, Oligocene, and Neogene, and why the volcanic output of the Izu and Marianas sectors diverged in the Neogene.

7. Data sharing and database development. Both MARGINS and JAMSTEC now have policies about sharing data in a timely fashion, and Workshop participants noted the need for placing data in publicly accessible formats and sites. This applies especially to routine marine geophysical information, but also to more processed geophysical and geochemical results. Progress on this topic may require policy level activity by organizations as well as enticements to individuals to undertake such projects.

D. Summary

As noted above and in the Subduction Factory Science Plan, the IBM arc has many features that allow fundamental subduction zone processes to be studied there. There are many respects in which it is becoming the world's best known oceanic subduction zone. In addition, its location is ideal for joint projects between US, Japanese, and other scientists. Three examples are noted at the end of Section B; more are included in the Topical Summaries. International cooperation is needed to mount the field, modeling, and experimental programs of the scale and resolution required for breakthrough science. The Workshop both illustrated and laid groundwork for further such cooperation, but it will require sustained bilateral initiative to carry it out. Periodic workshops on this type, postdocs and visiting fellows who move between institutions and countries, and joint field, numerical modeling, and experimental programs may be required. The advent of IODP may complement and accelerate such activity.

Izu Bonin Mariana 2002

http://www.ldeo.columbia.edu/margins/SubFac/IBM/IBM02.html

MARGINS Workshop on Subduction Zone Dynamics and Thermal Structure

Peter Van Keken[1], Scott King[2], and Simon Peacock[3]

1. Department of Geological Sciences, University of Michigan, Ann Arbor, MI, USA; E-mail: keken@umich.edu
2. Department of Earth and Atmospheric Sciences, Purdue University, West Lafayette, IN, USA
3. Department of Geological Sciences, Arizona State University, Tempe, AZ, USA

During October 4-6, approximately thirty researchers assembled at the University of Michigan for a workshop on subduction zone dynamics and thermal structure. The workshop was supported by the NSF MARGINS program and the Department of Geological Sciences at University of Michigan. The purpose of the workshop was to address some of the major tasks facing the modeling community including: benchmarking and evaluation of existing numerical approaches, developing strategies to exploit increased computer power and improved numerical techniques, and furthering the integration of numerical modeling into experimental and observational studies. Workshop participants included both those who primarily develop subduction zone thermal models and those whose expertise lies primarily in experimental or observational techniques but who are interested in subduction zone structure and processes.

The meeting was organized so that the majority of the available time could be devoted to group discussion. Discussion leaders were asked to provide introductory presentations to help stimulate the discussion. The participants seemed pleased with this format and there was certainly no shortage of discussion during the workshop.

The workshop began with an opening presentation by Bob Stern asking participants to consider whether it was the time to develop a community model for subduction zone thermal structure analogous to the global circulation models used in climate change research. The opening round of discussion focused on the pros and cons of a community model. The consensus that emerged is that there were many opportunities for the community to work more closely than they currently are; however, it was probably premature to focus community-wide effort on a single model. It was agreed that there is a need to benchmark current models and understand the strengths and limitations of the approaches currently being used. The idea that a set of community tools could be developed was discussed during the opening period and through out the workshop.

Next, each participant was invited to bring one overhead to the front of the room and describe their favorite subduction zone model. Aside from the discovery that some participants could not count to one, the variety of observations, models, and ideas presented helped to remind everyone of the diverse observations and inter-related problems that make subduction zone research challenging. Many of the overheads presented during the "my favorite subduction zone model" are available on the workshop website at:

http://www.geo.lsa.umich.edu/keken/subduction02.html

Observations

In the afternoon, workshop participants turned their attention specifically to identifying the key physical and chemical constraints on subduction zone processes, guided by Huw Davies presentation. While slab geometry, heatflow, gravity, and topography appear to be obvious constraints even these are not straightforward because of the load imposed by the volcanic arc. The geochemical constraints require accurate modeling of melt and fluid transport.

Continuing with the theme of observations, the workshop participants joined with the Michigan faculty to hear a lecture by Karen Fischer describing observations of seismic anisotropy in arc regions. Seismic anisotropy is an important observation because it can be used to map the geometry of wedge flow, assuming that the mechanism of crystal orientation is understood. At back-arc stations, the fast directions of anisotropy inferred from shear-wave splitting show strong variations between subduction zones. In the Tonga back-arc, fast directions are roughly parallel to the azimuth of subducting plate motion, while in the southern Kurils, fast directions lie parallel to a back-arc strike-slip shear zone. In both cases, the observed anisotropy can be modeled by lattice preferred orientation of olivine in simple flow models driven by coupling to observed three-dimensional plate motions, assuming that olivine a-axes align with flow direction or maximum finite extension.

Connection

On Saturday morning, the group reassembled and discussed the connections between this and other workshops. There were informal reports from the recent NSF-IFREE MARGINS Workshop on the Izu-Bonin-Mariana Subduction System. Bill Lehman advertised an opportunity to participate in a "State of the Arc" workshop next summer to be held in Oregon and Peter van Keken and Scott King drew attention to a subduction zone dynamics and thermal structure session planned for the EGS-EUG-AGU Joint Assembly in Nice France next spring. Thorsten Becker described an effort to benchmark codes that developed out of a European workshop and extended an invitation to participate the effort from the organizers of that project. The details of that effort can be found at:

http://www.geology.ethz.ch/sgt/Geobench/

There was a general consensus that the rediscovery of the effect of temperature-dependent viscosity on the thermal structure of the wedge, presented by several research groups, has been an important development in subduction zone thermal modeling efforts. The revived interest in temperature-dependent viscosity models could be traced back to discussions at the Margins Subduction Factory TEI held in Eugene Oregon in 2000.

Exchange of Ideas

The participants then divided into two groups, with the charge of answering: "what would observational and experimental researchers like to see from thermal models?" and "what would thermal modelers like to see from observational and experimental studies?" After an hour of brainstorming, the two groups came back together to exchange answers.

The 'observational' group came back with the following list of key observations: heatflow, gravity and geoid, topography and slab geometry (slab dip). There was a desire to see more slab thermal modeling studies report how well (or poorly) the models match these key observations. Other observations that the group would like to see slab modeling assess include: the distribution of volcanoes above slab in 2D and along-strike in 3D, the pressure and temperature conditions at the top of the mantle wedge, the low seismic velocities at the top of the slab, thickness variations observed in the upper plate, general seismic attenuation and velocity structure of the wedge, the presence or absence, and time duration of back arc basins, the asymmetry of Pacific subduction zones, limits of shallow seismogenic zone, seismic anisotropy of the wedge, and major element compositions of primitive lavas.

The modeling group addressed the strengths and weaknesses of the current modeling apporoachs. The arc-side boundary condition and the slip/no-slip transition along the fault were recognized as critical boundary conditions that need to be addressed. There are unanswered questions of how to treat elastic energy dissipation in dynamical models and a general recognition that dynamics are very sensitive to initial conditions when the slab and plate are driven by buoyant flow and not imposed kinematically. Time-dependent effects are not captured by most existing kinematic-dynamical models. To some extent, the subduction zone we observe today reflects its geologic and tectonic history. Modern observations of subduction zones only provide us with snap shots of a process that may vary significantly with time. Marc Spiegelman noted that most of the kinematic models and even some of the dynamic models do not produce accurate pressure fields and that this presents a major problem for combining fluid or melt transport with existing modeling approaches.

Figure 1. Participants in the Subduction Dynamics workshop. Bob Stern and others (top left) listen to Huw Davies (right). Simon Peacock faces the audience (bottom left).

Advances

After lunch the discussion focused on what was needed to advance beyond the current state, guided by Shun Karato and Marc Spiegelman. A consensus emerged that our limited understanding of rheology, including the effects of water and melt, presented a problem for furthering dynamical models. In addition, the transport and distribution of water appears to be a recurring theme with questions such as: what is the mantle wedge flow pattern, do subducting slabs contain a significant fraction of water in serpentine, what is distribution of water in wedge, does the basalt-eclogite transformation occur at equilibrium, and do dehydration reactions trigger earthquakes?

Then there was a long awaited discussion where some of the nitty-gritty details of computational modeling were discussed. A particular concern raised was the difficulty of modeling the dynamics of subduction zone formation given the limited understanding of the governing equations and expense of computational methods for modeling a

(continues on page 38)

Data Management in the Earth Sciences

Data management issues within the earth sciences have been a focus for intense work during the past few years. Important decisions and recommendations for the future of databases and data management in marine geology and geophysics were made at a workshop in La Jolla, California, in May of 2001 (see **http://humm.whoi.edu/DBMWorkshop/**). About a year ago, MARGINS adopted a policy for the management of data generated within the program (available for download as a PDF file at: **http://www.ldeo.columbia.edu/margins/PDF/MARGINS_Data_Policy.pdf**), and earlier this year, issued an announcement soliciting formation of data management systems in support of Focus Sites to meet the immediate needs and goals of the MARGINS community. The MARGINS Office is therefore motivated to summarize the progress on data management effort related to MARGINS. This issue of the MARGINS Newsletter is devoted to the "Data Management in the Earth Sciences", and contains the following articles:

New NSF-OCE Data Policy

Davip Epp

Division of Ocean Sciences, National Science Foundation, 4201 Wilson Boulevard, Arlington, Virginia 22230, USA; E-mail: depp@nsf.gov

The Division of Ocean Sciences is in the final stages of rewriting its data policy. The policy was last updated in 1994 and this revision will provide for a more systematic enforcement of the policy.

Investigators should first be aware of new requirements in the contents of proposals. The new Grant Proposal Guide (NSF 03-2) now requires that in the project description investigators, where appropriate, describe plans for preservation, documentation, and sharing of data, samples, physical collections, curriculum materials and other related research and education products. We expect that enforcement of this requirement will become more rigorous in the future.

The basic philosophy of the new OCE data policy comes from the NSF Grant Conditions and states that "NSF expects significant findings from research and education activities it supports to be promptly submitted for publication, with authorship that accurately reflects the contributions of those involved. It expects investigators to share with other researchers, at no more than incremental costs and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work. It also encourages grantees to share software and inventions or otherwise act to make the innovations they embody widely useful and usable."

The new policy maintains the sixty days requirement for submitting metadata to national archives and the two-year requirement for submitting the data. Where appropriate program-supported databases exist, PIs can satisfy these requirements by submitting to the program-supported database. It is expected that the program-supported data centers will be responsible for subsequently submitting appropriate data to the national centers.

In cases where no appropriate national data or sample repository exists, PIs are still required to detail how they will maintain the samples and/or data and how they will make it available.

The most significant change in the new policy is that PIs are now requiredto document in their final reports how they have complied with the data policy. Approval of a final report will depend on such compliance and failure to document it could delay the processing of any future financial support for the responsible Principal Investigator.

We recognize that complying with these requirement will entail some effort and expense on the part of the PI, and we expect PIs to budget for these expenses when writing proposals. We also recognize that archiving data and samples implies future use, and we encourage proposals to use archived data and samples.

WDC-MARE / PANGEA: a convenient RDBMS partner with MARGINS?

Nicolas Dittert[1], Michael Diepenbroek[2], Hannes Grobe[3], Reiner Schlitzer[3], Rainer Sieger[3], and Gerold Wefer[3]

1. Laboratoire des Sciences de l'Environment Marin (LEMAR-IUEM), Plouzané, France; E-mail: ndittert@wdc-mare.org
2. Centre for Marine Environmental Sciences (MARUM), Bremen, Germany
3. Alfred Wegener Institute Foundation for Polar and Marine Research, Bremerhaven, Germany

THE CHALLENGE OF MUTUAL SCIENTIFIC, SOCIAL, AND ECONOMIC BENEFIT

MARGINS has defined four focus initiatives, each with two focus sites, whose selection criteria are summarized in the respective science plans, cf. **www.ldeo.columbia.edu/margins**: Rupturing Continental Lithosphere (RCL), Seismogenic Zone (SEIZE), Source to Sink (S2S), and Subduction Factory (SF) initiatives. It is self-evident that any initiative comprises its own scientific approach. While one initiative will test hypotheses through the use of a suite of inter-connectable numerical and physical-process models with shifting boundaries, another initiative will rather proceed by field-, in situ-, laboratory- and experimental efforts. The diverse measurements will then be integrated at every level with geodynamic, physical or chemical models whose predictions will guide future data acquisition/drilling efforts. Eventually, the data collected will provide constraints for the next model generation, and modeling and observation will complement and stimulate each other.

In this specific project related environment data management merits a central position since it is responsible for archiving and documenting any scientific information beforehand, during and after a project funding period, and is carefully acting as mediator.

For some good reasons, there are quite a lot of scientific and economic initiatives engaged in continental (sensu Liu et al. 2000) and oceanic margins in general and in "MARGINS" margins in particular. To give a few examples: (1) Earthscope (a NSF-requested facility for the NSF Earth Sciences programme) shall improve the understanding of the evolution and dynamics of a continent; (2) NATMAP Shield Margin Project area studies a mineral rich portion of the Canadian Shield; (3) RIDGE 2000 (R2k) is a community-based science initiative focused on integrated geological and biological studies of the Earth-encircling mid-ocean ridge system; (4) UK and European Earth scientific communities have organized themselves to solve first-order problems considered important for both academia and the oil industry (e.g., Ocean Margins LINK; EuroMARGINS; InterMARGINS programmes). Other margin related research programmes investigate the complex topic of Global Climate Change (e.g., Continental Margin Task Team, CMTT; ex-NSF Land Margin Ecosystem Research, LMER; Research Center Ocean Margins, RCOM) rather than Geophysics.

Interdisciplinary, process-orientated studies and multinational co-operation between researchers involved in field data collection, numerical simulation and laboratory analysis promise unparalleled synergism necessary to understand the given complex, natural (i.e. biological, chemical, physical) systems and are supposed to cause substantial economic, scientific, and societal impact.

Exactly this kind of framework is manifesting the eminent role that data management merits if it succeeds to weave a web of liabilities between all levels. In such a scenario, scientific data management will play an outstanding, yet pivotal role by acting as a moderator between science and economy in trying to open up multi-national interchange and facilitate the release of data sets.

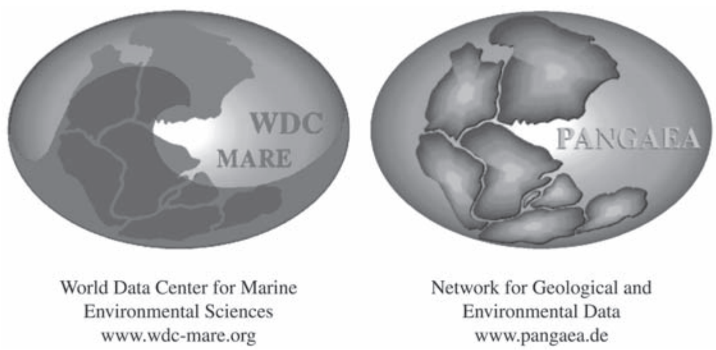

Figure 1. World Data Center for Marine Environmental Sciences (WDC-MARE; www.wdc-mare.org) and Network for Geological and Environmental Data (PANGAEA; www.pangae.de): An information management system and its operation platform.

DATA MANAGEMENT, INFORMATION MANAGEMENT — CAPTURING POLE POSITION?

Data management in marine Geosciences evolved from descriptive cataloguing to the relational digital record spanning a period of about one century: After the famous British H.M.S. Challenger deep-sea expedition (1872-1876) returned, an international team of investigators analyzed the staggering body of observations, converted them into records of qualitative and quantitative data, to eventually publish them as a 50-volume report (e.g., Murray and Renard 1891). Since the middle of the last century, inconceivable huge and heterogeneous numeric data loads came up during marine large scale projects (e.g., DSDP, JGOFS, ODP, WOCE). At this time, a data management strategy termed The-box-of-floppies-approach was developed. Data sets were supplied to the data center as discrete entities (usually on floppy disks) where they were checked, catalogued and stored on tape. On demand, clients were supplied with the data sets necessary to satisfy their requirements. Data management philosophy in this scenario was firmly focused on data archival (e.g., Lowry 2000). Today, the challenge of data management is to provide standardized import and export routines to support the scientific community with comfortable and uniform retrieval functions and efficient tools for the graphical visualization of their analytical and metadata (e.g., Dittert et al. 2002).

With respect to the ownership of extensive Earth scientific data collections, there has been a steady but irreversible shift over the last few decades, too. Until some fifty years ago, it was the academic institutions (e.g., Lamont, Scripps) that were the custodians of worldwide data sets. The rapid growth of the oil industry, efforts in global drilling, laboratory experiments, advances in n-D reflection seismic acquisition and processing, [...], have resulted in the generation of immense data bases whose ownership rests with the oil industry.

Where traditional funding organizations are not (yet) in a position to compete with industry in terms of data quality, coverage and rate of data acquisition, scientific initiatives such as MARGINS are forced to search mutual interest between industry and university. Ideally, any of the “initiators” would walk around, look out for obvious symbiotic relationships, and exploit them in order to minimize expenditure and to maximize profit for all parties. Alas, we know that interests are rarely common, and exploitation is not always multi-lateral. However, one of MARGINS’ challenges is to take advantage of any “initiative” during the execution of the program. Since the time frame is in the order of a decade and less, efficiency is supposed to be high, and milestones reachable quickly.

One logical step forward is the implementation of an information system that eventually would serve as a pivotal platform in order to transfer knowledge. Ideally in such an environment, one would choose for a data management system that (1) offers comprehensible scientific organization; (2) is entirely available on-line; (3) ensures reliable data import; (4) permits straightforward data retrieval/export; (5) provides software tools for data visualization and analysis; (6) sensitively supervises data policy during and after the funding period of any (housed and invited) of the initiatives; (7) operates on a longevity basis; (8) ensures at least basic financial autonomy.

MARGINS KEY DATA BASE MANAGEMENT QUESTION(S)

Fortunately, the basic allocation of data management tasks is already explicitly stressed out by MARGINS' data policy, has been adopted from the NSF, and precisely states what both parties (scientific community and NSF) expect from the eventual Data Management System (DMS) and Data Management Center (DMC) for the various MARGINS Focus Sites. Moreover, there are the Science Plans for MARGINS’ Focus Sites that unambiguously specify what kind of research is scheduled.

A short outlook would summarize as follows:

1. Any information that describes analytical data (e.g., parameter, method) or that describes general knowledge, i.e. metadata (e.g., project facts, cruise mnemonics, official station lists), shall be centralized in metadata catalogues that point to respective data centers.

2. Success is defined (inter alia) through formalizing and rapid release, sharing, and online access/dissemination of high quality data to maximize synergies, transfer, and integration between the partners and minimizing energy loss.

3. A straightforward list of variables/parameters comprises five categories, using a broad variety of temporal resolution from the Recent to the Past including time-series, and employing simple to sophisticated techniques:
   - In-situ and field observations
   - Laboratory experiments
   - Geo-dynamic and physico-chemical modeling
   - Monitoring efforts
   - Proxy parameter development

4. Steering committees shall foster coordination and symbiotic relationships between projects to ensure quality control of original, processed, derived, and interpreted data at all steps from acquisition through archiving, data exchange formats, metadata schemata, file structures, and sharing of standardized templates, tools, and procedures through Internet.

Key recommendations describe further aspects that should be taken into consideration. Different scientific communities already have organized their data through well-managed data collections that became standard for their discipline (e.g., IGPP, IRIS, NGDC, UNAVCO, ex-WDC-A). Other data, typically including results from “prior” research, are scattered among individual scientists because no standard archive exists. Yet, MARGINS’ multidisciplinary team approach demands open databanks that are properly supervised. Even if two types of databases are envisioned (i.e. one for each of the focus areas, and one for global data), scientific relational data base management systems (RDBMSs) are a priori well qualified to manage all data simultaneously. However, as any printed book is available in numerous libraries at the same time, no one would seriously argue against data (mind: identical data) mirrored at several suitable places, cf. ODP:

www-odp.tamu.edu/mirrors.html

In any case, the logical step is to employ an existing web-based front-end and SQL-based (RDBMS) backend information system to interrogate and manipulate data through Internet.

LOOKING INTO THE DATA MANAGEMENT MERITS OF WDC-MARE/PANGAEA

While the World Data Center for Marine Environmental Sciences, WDC-MARE, is considered as an information management tool, PANGAEA (Network for Geological and Environmental Data) works as the operating platform for WDC-MARE and represents the European standard of Spatial Data Infrastructures (SDI) (Fig.1). Due to a highly diversified affinity for data management within the scientific community, a dynamical approach is chosen that allows for a maximum of data with a minimum of a scientist's exhaustion. Ergo, data can be readily shared between participants, which is a relatively straightforward consideration and meets the principles of any modern data and information system plan (e.g., Boudreau et al. 1996). Historically, WDC-MARE/PANGAEA is deep-rooted in the development of a coordinated data management strategy for digital data of the Marine Geology and Geophysics community and in the identification of key data management problems facing this community.

PANGAEA is a scientific RDBMS and based on a three-tiered client/server architecture (Fig. 2, Eckerson 1995). On the client side (frontend - first tier) a number of interfaces are offered for access to the system. The middleware (second tier) is an application server with several components for the import, retrieval and maintenance of the data. A relational database finally is the central storage facility for all types of data (backend - third tier). Each of the client programs and the SQL server communicate with the intermediate layer, the application server, whereby the communication between application server and web clients requires an additional web server. All components are encapsulated as far as possible and use standard interfaces to communicate with each other. Middleware and frontend components are generic to ensure flexible functionality. Because open system design components can be altered or new ones added without affecting existing components this concept extends the lifetime of components and reduces overall maintenance costs.

WDC-MARE is operating under the auspices of the International Council for Scienceís (www.icsu.org) World Data Center system that comprises a network of some 50 WDCs covering any aspect in Geosciences (from A...Airglow, Mitaka, Japan to Z...Sunspot index, Brussels, Belgium). Its main missions constitute data/project management, data compiling/standardizing/homogenizing/archiving even for and from third parties, sample administration, information exchange, data publishing, site-mirroring (e.g., ODP) etc. WDC-MARE is accessible online via Internet through any connection and protocol, from any computer system that supports standard browser software, all over the planet, and at any time.

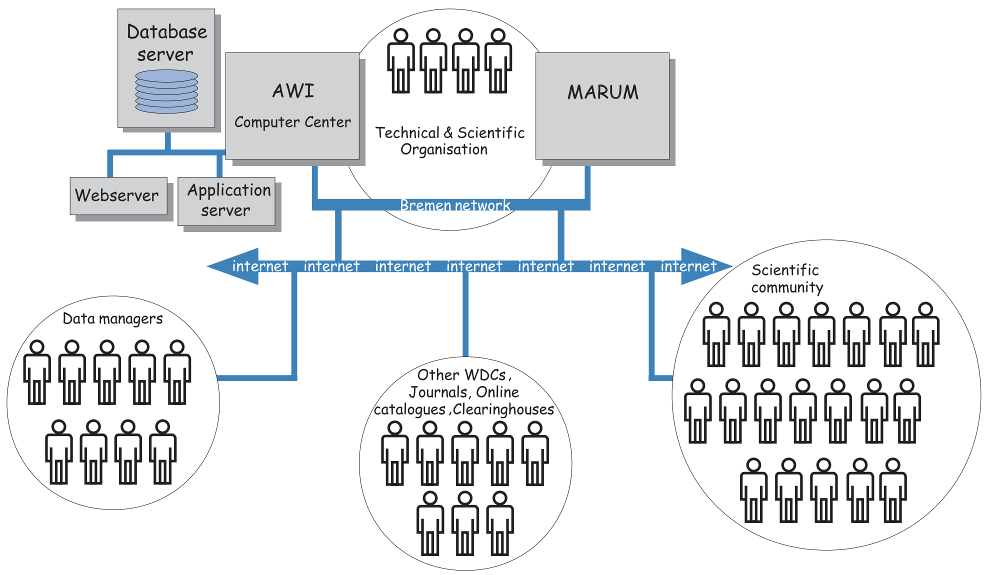

Figure 2. Institutional, technical, and scientific framework of PANGAEA and WDC-MARE. AWI computer center is maintaining all basic IT services to run the information system (i.e., hard- and software, local networks, backup services). Access to the WDC-MARE/PANGAEA servers is enabled by the Bremen high speed network (152 Mbit). Actually, WDC-MARE/PANGAEA is an active partner in 39 projects of manifold scientific impact and funding (local-regional-national-international-European-extra European). Access frequency to the PANGAEA servers doubles about every year.

Due to its standardized international information infrastructure, WDC-MARE/PANGAEA serves as backbone for the scientific community in marine and terrestrial Bio-, Chemical-, Geo-Sciences, working on any time frame (Recent - Past) and size of data unit (single scientist spread sheet - global data collections). On demand, working contexts can be elaborated comprising DBM solutions on-site (i.e., analyzing or visualization software or custom software solutions for research vessels, institutes, etc.), training units (single scientist - Masters courses), professional information management, etc. A highly flexible data model reflects today's standard for the analysis of complex data inventories or data mining (e.g., Han and Kamber 2000). Due to its leading front of presently available scientific information systems, WDC-MARE/PANGAEA gains pivotal functions beforehand, during and after the funding period of a project. Moreover, using WDC-MARE/PANGAEA becomes more and more good scientific practice.

Data preparation. Personal contact between data management staff and PI proves to be beneficial to both parties. Together with the PI, all important meta-information are collated: Project facts, cruise mnemonics, official station lists as well as notes on institutions and co-workers involved. The most critical performance is summing up and discussing the analytical data, which would serve as a first quality check: Are all values valid? Are all units consistent? Are all parameters adjusted? Are ridiculous and suspect numbers to be corrected or to be flagged or to be wiped out, respectively? Are methods checked for completeness? Are publications referenced? Finally, meta-information and analytical data both are converted into tab-delimited ASCII spread sheets following standardized input forms available online.

Data archiving. WDC-MARE staff at MARUM or data curators who are employed world wide for specific projects are responsible to archive analytical data and the corresponding meta-information. They use a highly sophisticated client/server data management tool, 4th DimensionÆ, that permits direct access to the central data base server via Internet. PANGAEAs formal RDBMS standard, the so-called data base model (Fig.3), ensures maximum data security and quality. To upload data collections, any PANGAEA client offers import routines for CAMPAIGN, SITE/EVENT, SAMPLE, REFERENCE, PARAMETER, ARCHIVE, and analytical DATA. These files are loaded online into a temporary table at the server side where a second quality check is initiated and meta-information are examined for coherence with entries already archived: Are station lists identical to the existing catalogue? Are geographical positions accurate? Then, analytical data are compared to pre-defined settings: Do there exist duplicates? Are the numbers situated between upper and lower ranges? Once a data series has cleared this hurdle, data are transferred to the import server and stored temporarily in order to update the data warehouse at regular intervals. Finally arrived at the export server, a uniform resource locator (URL) is generated and data are offered to the scientific community.

Data retrieval. There exist three different ways to retrieve data from WDC-MARE, which may be distinguished by the skill that is required from the user. In any case, the general numeric output from the database are tab-delimited ASCII files that accommodate both analytical data and meta-information. All data can then be further processed in using commercial software. A help system and tutorial are provided for online support. (1) PangaVista is a search engine similar to AltaVista. Complementary to PangaVista is the download of lists that were compiled to portray whole projects, or institutes, or publishers, respectively. (2) Some more skill is required if the database is accessed by the data mining tool ART (Advanced Retrieval Tool). (3) The direct access to the server, however, is the only data processing tool that permits both reading and writing access. Since this sophisticated application requires a particular measure of expertise, the gateway is limited to the data management staff.

PangaVista. This simple web-based search engine enables the retrieval of whole data sets referenced by dynamic URLs. PangaVista is linked with a Thesaurus comprising any meta-information related to the analytical data and thus can be used with a variety of keywords (e.g., principal investigator, author, title, parameter, method, project, sampling location, etc.) Keywords can be combined to create boolean expressions, with syntax identical to that used by AltaVista. Results contain a short description of the retrieved data sets with the URLs to the related data. Data can be downloaded as HTML or tab-delimited text. All data sets contain a so-called metadata header according to the data interchange format (DIF) standard used by the NASA's Global Change Master Directory (GCMD).

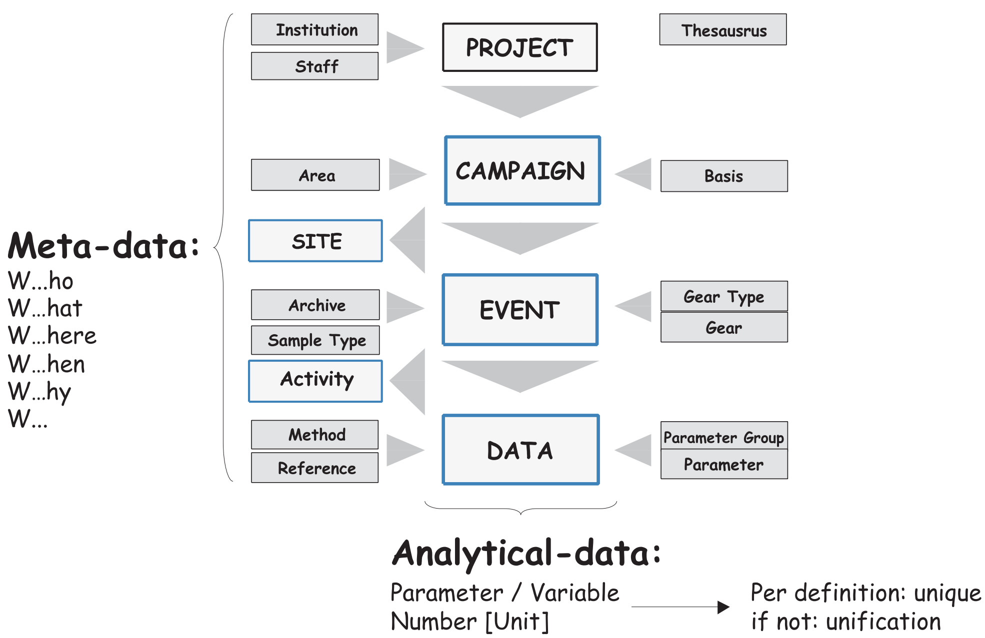

Figure 3. WDC-MARE/PANGAEAs highly flexible data model comprising web-based online access to heterogeneous and dynamic metadata and geo-coded analytical data. This is achieved by a combination of a simple and fully normalized RDBMS structure in combination with the functionality of middleware components. The data model reflects todayís International spatial data infrastructure and metadata standards and serves for analysis of complex data inventories or data mining (Han and Kamber 2000).

ART is a web-based tool to retrieve all types of analytical data and meta-information. The flexibility of this tool allows for the definition of complex retrievals as well as for the individual configuration and formatting of result sets that are subsequently visualized as listings, plots, and/or maps. The simplified data structure of PANGAEA is the opening user interface that allows users to enter all levels by selecting the desired field. Results can be downloaded as tab-delimited text, or as plots and maps in '.ebf' or '.ps' format. Through its functionality ART can be classified as a data mining tool. It enables the user to retrieve analytical data for any section of the globe and to combine any type of data, provided that the geo-coding of the data is compatible. For the graphic user interface (GUI) of ART a Java applet is employed; the middleware transforms the logical requests and prepares the result sets for the client. A context sensitive help system is supplied for all levels.

Data quality. Validation and verification of data are the two substantial aspects during data archiving: External discovery of errors will do more to destroy the credibility of the database and the data management group than anything else. However, the definition of what is correct is far from straightforward. What is correct can quite often be a matter of opinion and opinions are subject to change as scientific knowledge (Lowry and Loch 1995). While it is not essential to have only excellent data sets archived it is, however, indispensable that exact information on the quality is provided. Complete meta-information, including, in particular, analytical method and reference where the data were published first is crucial to depict any data set. Manual quality checks are supplemented by an evolving system of generic and parameter-specific validation routines that base on the definition of parameters and analytical methods which require standard unit, upper and lower ranges, precision and tolerance of the analytical value. Data collections from third parties must be treated even more carefully since data are trawled from different sources, each with their own quality standards. In any case data access is controlled by the data owner or Principal Investigator, respectively, since WDC-MARE archives published as well as non-public data.

Data visualization and analysis software. PANGAEA supplies any user with a series of freeware/software products that are well-established scientific tools, of excellent quality, and used worldwide. Ocean Data View, PanMap, and PanPlot are built to be used as standalone-applications on the scientists computer or online in conjunction with browser software connecting to the information system. These interactive software modules support direct access to any export format of PANGAEA. Maps, plots, and cross section profiles can be exported in platform specific interchange format and further processed in using commercial software. Since they belong to the operating system PANGAEA rather than to the information system WDC-MARE they are explained in detail elsewhere (Diepenbroek et al. 1999; Diepenbroek et al. 2002).

Ocean Data View (ODV). ODV freeware/software package is employed for the interactive exploration and graphical display of multi-parameter profile or sequence data. Although originally developed for oceanographic observations only, the underlying concept is more general, and data or model output from other disciplines (e.g., Geophysics) can be maintained and explored as well. ODV data format is designed for dense storage and direct data access, and allows the construction of very large data sets, even on affordable and portable hardware. ODV supports simultaneous display of original and "interpreted" data by color separation. Two fast variable-resolution gridding algorithms allow color shading and contouring of gridded fields along sections and on general 3D surfaces. ODV runs on PCs under Windows and on UNIX workstations under SUN Solaris. Software and extensive sets of coastline, topography, river-, lake- and border outlines as well as various gazetteers of topographic features are available. A gallery of prepared plots of property distributions along World Ocean Circulation Experiment (WOCE) sections provides a quick overview over tracer fields in the ocean and, apart from the scientific use for oceanographic research, can serve as tutorial material for introductory or advanced courses on oceanography (cf. Brown 1998; Schlitzer 2002).

PanMap. For geographical presentation of data, the PANGAEA mapping tool PanMap was developed. This simplified Geographical Information System (GIS) manages stationary and vector data that organized in layers. The freeware/software allows for the configuration of maps with different projections, import of additional vector and stationary information, and labeling of stationary information. Styles of map elements (e.g., color scales) can be changed for stationary data. An extensive collection of map resources is available online, among them the General Bathymetric Chart of the Oceans (GEBCO), the World Vector Shoreline (WVS), and the Global 30 Arc-Second Elevation Data Set (GTOPO30). Tools are provided to digitize individual maps whose resulting output data or any individual contour lines can be converted into the PanMap format. This software is available for Macintosh and Windows platforms. PanMap has become one of the most frequently used freeware GIS worldwide.

PanPlot. PANGAEAs plotting tool PanPlot allows the visualization of multiple data series as 2-D plots or ternary diagrams, respectively. Up to 255 parameters can be plotted at uniform scales. PanPlot manages numerical, date/time, and text data. Values on the ordinate can be of numerical or date/time type. The input format is plain ASCII code/spreadsheet. Scales and graphic features can be modified, and distinct parameters can be selected from the data matrix. Labeling and whole plots can be rotated by 90 degrees. Data sets exported from PANGAEA are interpreted by PanPlot including the so-called meta-header, which is shown in a separate window. The software is available for Macintosh and Windows platforms.

Operation resources (Fig.2). The institutional, technical, and scientific framework of PANGAEA or WDC-MARE, respectively, is grant-aided by the Alfred Wegener Institute (AWI) and the Center for Marine Environmental Sciences (MARUM) on a long-term perspective. AWI computer center is maintaining all basic IT services to run the information system (i.e., hard- and software, local networks, backup services). Fast access to the PANGAEA servers is enabled by the Bremen high speed network (152 Mbit). Data management services on an international level are supplied since 1996. In October 2002, WDC-MARE/PANGAEA is an active partner in 39 projects of manifold scientific impact and funding (local-regional-national-international-European-extra European). The access frequency to the PANGAEA servers doubles about every year. Actually, the average number of visitor sessions per day amounts to 200, and the mean number of daily data base queries totals 250, alone some 60 being retrievals for analytical data. Access logs and the response of users show a growing significance and acceptance of Internet based scientific information systems.

CONCLUSION, PERSPECTIVE, AND RECOMMENDATION

MARGINS is setting an enormous standard with respect to its data policy that will not be easy to satisfy. Likewise, MARGINS' scientific approach requires a good availability of gigantic amounts of analytical data (in the order of TByte) that goes far beyond the capacities of conventional relational data base management systems (RDBMS).

However, the World Data Center for Marine Environmental Sciences in close cooperation with the Network for Geological and Environmental Data (WDC-MARE / PANGAEA) can meet quite a subset of the requested workload whose duties might include:

- Coordinated and appropriate, web-based data- and information management strategies under long term aspects.
- International spatial data infrastructure and metadata standards.
- Metadata catalogue management comprising description of analytical data (e.g., parameter, method) or general knowledge (e.g., project facts, cruise mnemonics, official station lists) with links/URLs to respective data centers.
- Analytical data management for any type of geo-referenced analytical data (In-situ-, field-, and monitoring observations), time series, and laboratory experiments through a highly flexible data model under any temporal resolution from the Recent to the Past (Fig.3).
- Formalizing and rapid release, sharing, and online access/dissemination of high quality analytical data (cf. Dittert et al. 2001).
- Synthesis of different data resources (e.g., JGOFS; **www.pangaea.de/Projects/JGOFS/data.html**).
- Information management between projects to ensure quality control of original, processed, derived, and interpreted data at all steps from acquisition through publication, data exchange formats, metadata schemata, file structures, and sharing of standardized templates, tools, and procedures through several Internet interfaces.
- A straightforward, dynamic, extendible list of some 15.000 variables/parameters.
- Proxy parameter development (cf. Fischer and Wefer 1999).
- High-quality freeware/software tools online (cf. **www.pangaea.de/Software/**).

In any case, WDC-MARE/PANGAEA would be pleased to offer its services to MARGINS as one of its several robust healthy partners.

REFERENCES

Boudreau PR, Geerders PJF, Pernetta JC (1996). LOICZ data and information system plan No. LOICZ/R&S/96-6, ii+62 pp.). LOICZ Office, Texel, The Netherlands

Brown M (1998) Ocean Data View 4.0. Oceanography 11(2):19-21

Diepenbroek M, Grobe H, Reinke M, Schindler U, Schlitzer R, Sieger R, Wefer G (2002) PANGAEA - an information system for environmental sciences. Computers & Geosciences 28(10):1201-1210

Diepenbroek M, Grobe H, Reinke M, Schlitzer R, Sieger R (1999). Data management of proxy parameters with PANGAEA. In: Fischer G, Wefer G (eds.): Use of proxies in paleoceanography: Examples from the South Atlantic. Berlin Heidelberg (Springer-Verlag): 715-727

Dittert N, Corrin L, Diepenbroek M, Grobe H, Heinze C, Ragueneau O (2002) Management of (pale-) oceanographic data sets using the PANGAEA information system: The SINOPS example. Computers & Geosciences28(7):789-798

Dittert N, Diepenbroek M, Grobe H (2001) Scientific data must be made available to all. Nature 414(6862):393

Eckerson WW (1995) Three tier client/server architecture: achieving scalability, performance, and efficiency in client/server applications. Open Information System 10,1 3(20):1-12

Fischer G, Wefer G (1999) Use of proxies in paleoceanography: Examples from the South Atlantic. Springer-Verlag, Berlin Heidelberg, 735 pp

Han J, Kamber M (2000) Data mining: Concepts and techniques. Morgan Kaufmann, San Francisco, CA, 500 pp

Liu K-K, Iseki K, Chao SY (2000). Continental margin carbon fluxes. In: Hanson RB, Ducklow HW, Field JG (eds.): The changing ocean carbon cycle. Cambridge (Cambridge University Press): 187-239

Lowry RK (2000) The BODC relational approach to the management of water bottle data. Oceanographic and Atmospheric Data Management:in press

Lowry RK, Loch SG (1995). Transfer and SERPLO: powerful data quality control tools developed by the British Oceanographic Data Centre. In: Giles JRA (ed.):, Geological data management. (Geological Society Special Publication): 109-115

Murray J, Renard AF (1891) Report on deep-sea deposits, based on the specimens collected during the voyage of H.M.S. Challenger. Her Majesty's Stationary Office, London, 525 pp

Schlitzer R (2002) Interactive analysis and visualization of geoscience data with Ocean Data View. Computers & Geosciences 28(10):1211-1218

MG&G Database Management Efforts at Lamont-Doherty Earth Observatory

Bill Haxby, Bob Arko, Suzanne Carbotte, Dale Chayes, Suzanne O'Hara, and Bill Ryan

Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, New York 10964, USA; E-mail: bill@ldeo.columbia.edu

Lamont's Web-based MG&G database efforts began in the early 1990s with the RIDGE Multibeam Synthesis Project. In recent years, this has grown to encompass a wide variety of data sets, including underway and borehole geophysics, multichannel seismics, core descriptions, and geochemical analyses. Our data management efforts are mainly directed toward the development of tools for efficiently exploring databases and the standardization and management of metadata.

An important focus of our data management effort is the development of MapApp, a Java™ applet (also a standalone application) that provides capability to interactively explore and visualize diverse data sets in a map-based point-and-click interface. Our goal is to offer access to all data in our databases through an interface that is as efficient and intuitive as possible, and can satisfy users ranging from the professional scientist with specific data needs to the curious surfer who just wants to "see what's out there." To accomplish this, we provide search capability through a navigable map with selectable overlays, menu options and selection dialogs. We are also developing a set of analysis and visualization tools for inspecting and interpreting data.

MapApp was initially developed to provide easy access to high-resolution multibeam maps and digital elevation models (DEMs) for the NSF funded RIDGE Multibeam Synthesis. Its central component is a global physiographic map that has zoom in/out and pan capabilities, and upon which data locations may be displayed. Other visual components of MapApp include selection dialogs, tables, X-Y graphs, and a viewer for seismic images. The multibeam interface includes: identification of high-resolution surveys by toggling a mask at any map scale; loading map images at the current viewing resolution, down to 200 m pixel size; and loading DEMs that may be contoured or used to select and display a depth profile. Interaction with other databases may be initiated by selecting items from a menu. Typically, a table will appear with metadata information for the selected database, and symbols representing stations or tracklines will appear on the map. Items may then be selected from the table or map (by clicking a symbol) for further analysis. For example (see figure), double clicking on a ship track brings up a window with profiles of the depth, gravity anomaly and magnetic

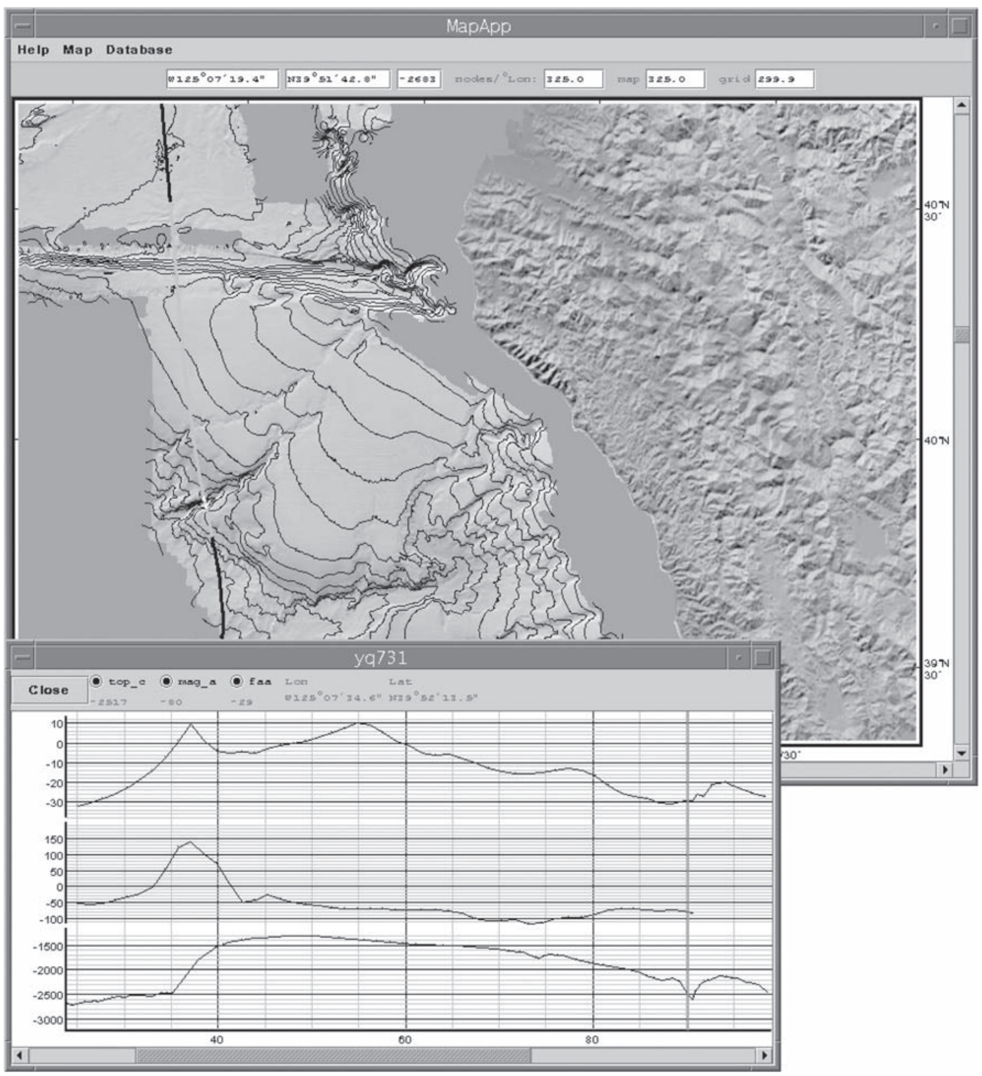

Figure 1. MapApp's interface for the LDEO underway geophysics database. The map window displays high-resolution multibeam bathymetry off of N. California, which has been contoured, after loading a DEM, at 200 m. The profiles window shows depth, magnetic anomaly and free-air gravity along the highlighted portion of the ship track. The cursor position in the profile window is also highlighted on the map.

anomaly. The profiles are scrollable and scaleable, and the cursor location in the profile window is continuously monitored, so that measured values are displayed in a text window, and the location of the point is shown on the map.

Database interfaces have been developed, or are actively being developed, for: The RIDGE Multibeam Synthesis; the LDEO underway geophysics database; the LDEO Borehole Research Group's ODP well-logging database; the RIDGE Petrology database (PetDB); multichannel seismic and sonobuoy databases; LDEOs core repository database; and ocean heat flow measurements.

Current development efforts are concentrating on standardizing the structure of database components (selection dialogs, treatment of metadata), providing download capability for various data types (grids, images, tables) in a variety of standard formats, enhancing the existing analysis and visualization tools, and adding new tools and database modules.

In the near future, MapApp will become a portal for exploring on- and offshore geophysical data in the South Polar Region, as we begin a major Antarctic database effort with NSF/OPP (Office of Polar Programs) funding.

A second major focus of our data management effort is the development of a more robust framework for managing our metadata. As our data sets grow more diverse, and funding agencies increasingly require prompt release of data into the public domain, it is imperative that we document our data sets as efficiently and thoroughly as possible. We have adopted the Directory Interchange Format (DIF), a subset of the Federal Geographic Data Committee (FGDC) metadata standard, which is widely used by U.S. agencies and many other countries. The DIF standard documents a data set's origin, type, location (in time and space), resolution, personnel, sensors, observables, references, citations, and history in a simple text format. Controlled vocabularies exist for key fields, which are moderated by a standards team at NASA and are expanded continuously to include new disciplines and data types.

In addition to making metadata available via MapApp and other public Web interfaces, we are also developing a dedicated server based on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and the Extended Markup Language (XML). OAI is rapidly emerging as the global standard for public exchange of metadata, and has been adopted by major U.S. digital library initiatives such as the National SMETE Digital Library (NSDL) and the Digital Library for Earth Science Education (DLESE). By adopting this standard, we immediately expose our metadata to a far greater public audience.

NSF-MARGINS Theoretical Institute on the Seismogenic Zone Experiment (SEIZE)

Snowbird, Utah, March 16-21, 2003

Processes occurring on the shallow seismogenic interface between subducting and overriding plates are responsible for generating the world's largest earthquakes and tsunamis, and are the focus of significant efforts in the fields of seismology, geodesy, tectonics, and natural hazards. These processes are the focus of the SEIZE (*SEI*smogenic *Z*one *E*xperiment) initiative of NSF's MARGINS program.

A five-day workshop will be held in Snowbird, Utah to review our current understanding, and plan for further studies of the Seismogenic Zone. Central America and the Nankai Trough, Japan, are two focus areas for SEIZE (the Science Plan is available on the SEIZE web pages at the MARGINS web site: http://www.ldeo.columbia.edu/margins). While the meeting will provide current information and syntheses about the characteristics of Central America, the Nankai Trough and other subduction zones as they relate to themes in the SEIZE science plan, a major emphasis will be to address more general and theoretical topics related to the subduction zone and the earthquake process, to synthesize outstanding problems, and to plan future experiments and collaborations. The theoretical institute format will be via a series of keynote presentations and impromptu talks and selected thematic poster sessions.

Participants will be chosen from applicants to this announcement. Selected applicants will be provided with full or partial funding of their costs for air travel, accommodation and meals. We encourage applications from all those interested in participating in this scientific endeavor, including those from outside the USA, Central America and Japan, and especially encourage applications from assistant professors, post-docs, and graduate students. Applicants should prepare a brief (no more than two pages) resume (CV), and a brief (half-page) statement of why they are interested in participating in the meeting and what they will contribute to it. Detailed instructions are included on the web application form.

Applications should be submitted via the web form at:

http://www.ldeo.columbia.edu/margins/SEIZE/ti03.html

Applications close on 13 December, 2002 with participant selection by 6 January, 2003.

Inquiries should be directed to the conveners:

Tim Dixon:	tdixon@rsmas.miami.edu
Eli Silver:	esilver@es.ucsc.edu
Kevin Brown:	kmbrown@ucsd.edu
Kevin Furlong:	kevin@tahiti.geodyn.psu.edu
Seth Stein:	seth@earth.nwu.edu
Casey Moore:	cmoore@earthsci.ucsc.edu

Seismic Reflection Data Access: The University of Texas Institute for Geophysics Experience

Thomas H. Shipley, and Lisa M. Gahagan

University of Texas, Institute for Geophysics, 4412 Spicewood Springs Road, Austin, TX 78759, USA; E-mail: tom@utig.ig.utexas.edu

Scientists have a responsibility to preserve observations fundamental to developing new ideas. The first choice, peer-reviewed publication, is not fully adequate for seismic data and thus, other efforts for preservation are required. The demand for academically-acquired seismic reflection data remains high because of their often-unique experimental design or location. Recognition of problems in finding existing data led a 1999 workshop to state that "improving accessibility to seismic reflection data might be the single most important component of any strategy to increase scientific returns from marine reflection activities" (**www.ldeo.columbia. edu/margins/SeisWorkshop.pdf**). To achieve this objective requires discovering the existence of data through some comprehensive search scheme and then obtaining access. For many scientists, access means a binary data file, but to others it means to an image; they are equally important. Satisfying basic access problems need not be difficult and would then expand the number of users, helping rebuild the marine seismic reflection discipline. The 2001 MG&G Data Management Workshop and its report provide a framework to launch new efforts in handling seismic data with the call for an overarching metadata management system fundamental for seamless basic standards and procedures across distributed data centers and data types (**humm. whoi .edu/DBMWorkshop/data_mgt_ report.hi.pdf**). MARGINS has issued an announcement soliciting formation of data management systems in support of focus sites to meet the immediate MARGINS needs and goals in the absence of a broader-based NSF plan. Experience gained in such efforts should help guide our scientific community in developing more comprehensive systems over the next decade.

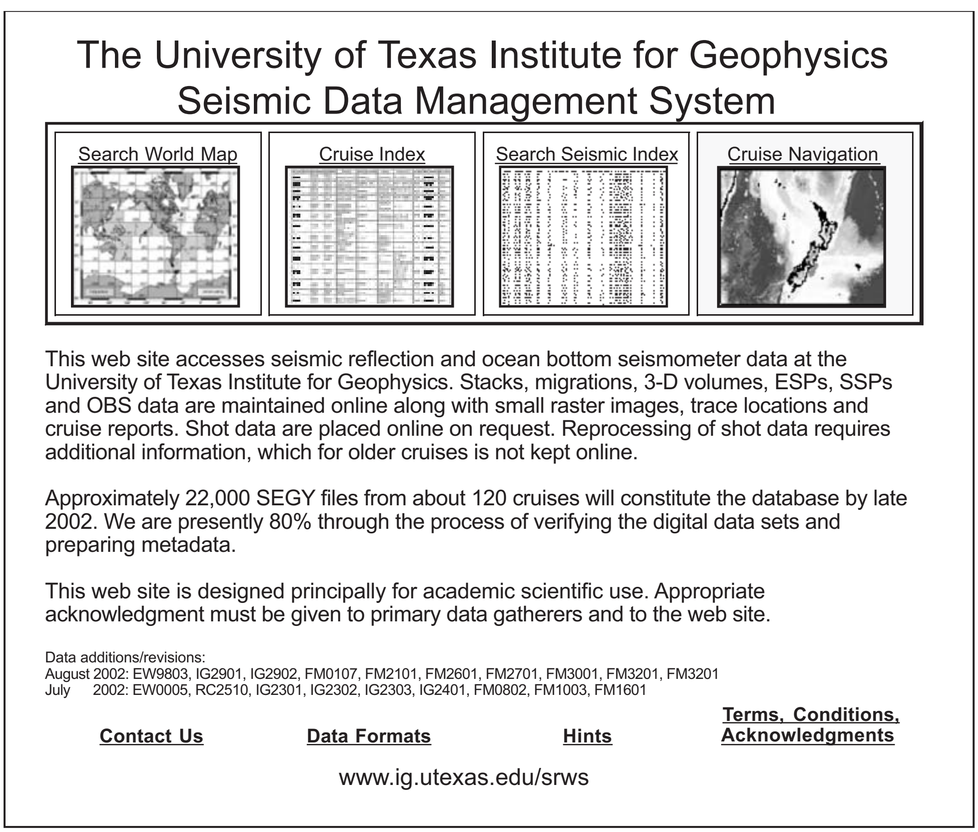

Figure 1. *Entry page to the UTIG Seismic Data Management System*

UTIG Seismic Data Management System

The Institute for Geophysics (UTIG) collected its first multichannel seismic data on January 11, 1974. Since then the seismic data archive has been maintained with support from an industrial associates program into the early 1980's, then by ever-reducing institutionally funded staffing levels into the late 1990's, and from direct support in seagoing science grants (mainly into the late 1980's). There has been a steady decline of support for these activities in part because of changes in operations (e.g., UTIG discontinued ship operations in 1988 and the order imposed by a centralized processing center was replaced by the anarchy of desktop processing in 1982). Over time we probably lost ~5% of the processed data and ~10% of the shot data. How these data were archived is not dissimilar to other major geophysical data collections held by oceanographic institutions.

By early next year we will have finished creating a searchable, web-based open-access metadata index of shot data, processed data, navigation, maps, seismic images and technical cruise reports for our entire collection. The scope of this project is to develop a prototype data management system (DMS) as a single access point to seismic data for UTIG users while providing an interim response to NSF guidelines for data sharing and archiving. Our focus has been on data recovery, identification, and metadata creation with relatively little effort on the web-access tools beyond static images and searchable SQL-based metadata. The web site accesses all seismic reflection and ocean bottom seismometer data housed at the Institute for Geophysics (~22,000 files from about 120 cruises). SEGY binary files of single-channel, stacks, migrations, 3-D volumes, ESPs, SSPs and OBS data are maintained online along with small raster images and trace locations. About 80% have been verified during our ongoing effort to create associated metadata. Shot data are placed online on request and have all been reformatted to SEGY (with the potential for some loss of extended header information, when present). Reprocessing shot data requires supporting information, which are mostly available offline for older data sets. A raided disk contains copies of all of the processed (derived) files while ~3 terabytes of mainly shot data are on tape, which provides long-term data integrity but needs migration and media replacement every three years. This is a relatively simple and inexpensive solution absent a more robust near-online solution for shot data, for which requests remain manageable.

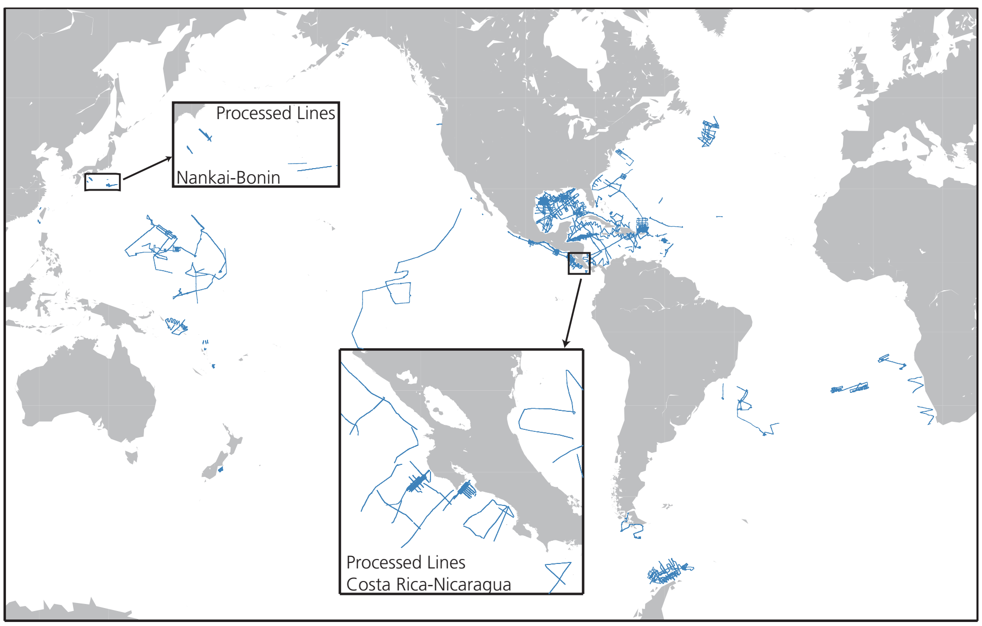

Figure 2. Seismic data in the UTIG SDMS as of September 1, 2002. Insets are derived (processed) data in three MARGINS focus sites.

Our efforts at acquiring permission to include data in the DMS reveal two significant concerns among data providers that have to be resolved for full participation. One is related to data sets that are not necessarily in the public domain (such as those funded by non-governmental programs, gifts, etc.) and another is a mixed set of issues related to potential collaboration, long-duration projects and the simple desire to know who is using the data. Part of the solution is to implement access restriction to some binary files and large-scale images. However, all metadata appear in the DMS. We presently distinguish three types of restriction. (1) Data are part of specially funded proprietary projects including many holdings in the Gulf of Mexico and Caribbean (~31 cruises). Usually, for academic investigators, individual lines can be released and data sets are available with UTIG scientific collaboration. Data are available for commercial applications with negotiated terms, conditions, and cost (a rare event). (2) Data are in use by active UTIG research programs (~7 cruises). (3) Data are processed profiles, representing the intellectual investment of investigators who still expect to use these data in future studies (~2 cruises). Data of type 2 and 3 often are available by request from the cognizant scientist. The second issue is related to wanting some recognition and to know how the data are being used. So we are planning to implement basic tracking of data downloads (with prior notification to the user) and reporting the activity to the data provider. We believe these access restrictions and usage information are necessary to gain the full cooperation of data holders, at least initially.

In January 2002 we reached a point where half of the processed data was online. Since then, on a monthly basis, web activity logs track about 500 different host machines and 12,000 hits of 3,500 different pages or files (excluding internal use). The most popular binary downloads are the Barbados 3-D volume, lines from South Atlantic African margin, Newfoundland Basin, and a regional line across the Blake-Bahamas margin (1975 vintage). Of the non-commercial, name-correlated domains, the big users are approximately: U.S. educational (25%), Italy (5%), Germany (5%), France (3%), and UK (2%). The remaining usage is by networks, commercial, non-profits, and those without a name service.

Holdings Relevant to MARGINS

The web site:

www.ig.utexas.edu/srws

is the portal for entry into the UTIG DMS. Below are a list of cruises in the database that may be relevant to MARGINS focus sites. Many of the data are relatively old but still valuable for regional settings. Others are so new that some access restrictions are still in place.

IBM Subduction Factory

Cruises: FM3507 (1987), FM3505 (1987). This was a project of B. Taylor and G. Moore and they are the primary data holders though the DMS has some shot data. We have some OBS and associated seismic reflection data for FM3507.

Nankai Trough Seismogenic Zone
Cruises: EW9907 (1999), FM3506 (1987). These cruises were both in the Nankai Trough region off Shikoku. The Ewing cruise is the 3-D program and the FM3506 was a two-ship MCS, ESP and SSP program. The Ewing shot data is available.

Nicaragua-Costa Rica Subduction Factory and Seismogenic Zone
Cruises: EW0005 (2000), EW9502 (1995), FM3502 (1987), TW8203 (1982), IG2903 (1978), IG2902 (1978), IG2407 (1977), IG2403 (1977), and IG2402 (1977). We are presently transferring the recent Ewing cruise to the system. All of the rest of the data are available.

New Zealand Source-to-Sink
Cruise: EW0001 (2000) Canterbury Basin high-resolution seismic survey.

Challenge of Capturing Processed (Derived) Data

There is understandable reluctance by some scientists to provide data and supporting information to any DMS. However, for field data it is possible to make this fairly easy and almost transparent to the scientist. Also for field data, time-based access restriction rules are relatively simple. The much more difficult part is capturing processed (e.g., stacked) data, which represent a far greater intellectual investment. The first issue that comes up is the time and effort required in locating files and gathering up information needed for creation of metadata, though this is not a valid excuse for relatively recent data. More fundamentally, the reticence is related to the intellectual investment and potential future use, such as representing a continuing body of observations being collected over a career or expectations of use by future students. In addition scientists desire recognition and want to promote collaborations, thus they like to know who is using data they collected or processed. Thus, for our DMS we are planning to implement methods to track data downloading and requests to lift data access restrictions and supplying this information to the scientists and DMS management. This will ease some of the misgivings scientists have in making derived data available to DMSs.

Instead of debating these data provider issues our approach is that data access restrictions will remain the sole preview of the scientist, with no exceptions. This seems fundamental to getting all the data and metadata into the searchable index, and without a complete index the value of the entire effort is greatly diminished. If a user can discover through the DMS that data exist, then it will be possible to find an agreeable solution for data release. The data management systems should not be setting access policy since it likely will be counterproductive to data submissions. An overriding objective should be to get as much data as possible into a DMS by expiration of the grant since later it becomes more difficult and costly.

Other Issues

There are many public domain databases of marine seismic data. Particularly groundbreaking was Antarctic seismic data library (walrus.wr.usgs.gov/sdls/contents.html) that codified submission policy and formats for data collected around the Antarctic and available offline. The U. S. Geological Survey has numerous other links to specific data sets that are mostly offline, as does the National Geophysical Data Center. Among the major oceanographic institutions, Scripps Institution of Oceanography lists 28 expeditions of shot data, some of which are online;

http://sioseis.ucsd.edu/reflection_archive/index.html

Many scientists also maintain personal archives and web presence.

There are numerous examples of project-based databases created by small groups of scientists. These tend to be purpose built, contain a broad range of vetted data (seismic, multibeam, plate models, geology, etc.) and are the primary resource used in day-to-day planning and research by many scientists. Data management systems should focus on making it easier for scientists to create these purpose-built databases. For instance a seismic DMS should include a single Internet pointer to comprehensive metadata and it should be possible to download both the seismic data and associated navigation translated into commonly used commercial (e.g., Landmark, Geoquest) and non-commercial software packages.

Even at the institutional or departmental level, capturing recent processed data will require significant peer-pressure. It is easy to say that there are requirements from funding agencies for data release, but the reality is more complex. For example the derived data is not a well-described entity. Is it a simple water velocity stack, a depth migration, or some other product? Often the best outcome is produced years after grant expiration. Specific mandates are not desirable since the derived products are completely arbitrary, as one should encourage in a curiosity driven endeavor. As scientists begin to see the value of such databases they likely will become more willing contributors. These challenges are not peculiar to UTIG; other institutions will encounter similar problems. An article by Helly, et al., 2002 (“Controlled Publication of Digital Scientific Data” in Communications of the ACM, 45, 97-101) discuss some of the issues related to tracking data retrieval and the need for notification of data providers.

Recent efforts to clarify data sharing policies in OCE and EAR are helpful and some data submission enforcement may be needed; though the data management systems should not be directly involved. A certain number of distributed sites will be necessary for application of peer-pressure and control issues associated with derived data. However, these need not be expensive undertakings once the underlying MG&G MMS defines the minimum data structure; the resulting metadata searches will be completely transparent to users. Our main purpose has been to catalog the extensive UTIG collection, simplify maintenance of the digital data and input data from the last decade. We have not concentrated on access tools since these are much more subjective and expensive to create.

We are appreciative of NSF funding to preserve the UTIG database but remain concerned about efforts to provide a more systematic approach to data archiving and access. We had hoped that the MG&G Data Management Workshop vision would have led to some specific NSF responses by now. In the interim MARGINS is developing data management systems for their program.

Ridge 2000 Integrated Study Site Data Policy, Data Management and Databases

Deborah R. Hassler
Ridge 2000 Coordinator
Department of Biology, Penn State University, University Park, PA 16802, USA; E-mail: dhassler@psu.edu

A critical aspect for success of the Ridge 2000 program is the open sharing of data. For Ridge 2000 to succeed as a community program, all potential investigators must have ready access to data from the Integrated Study Sites to compete equally with previous investigators at a given site. The program is committed to facilitating equitable access to all data that can be used to develop hypotheses and research proposals. To this end, the Ridge 2000 community has adopted a data policy. In accepting NSF support within the Ridge 2000 program, each principal investigator is obliged to meet the data policy requirements as an integral aspect of their participation in the program. Rapid dissemination of metadata and data will maximize technology transfer across the program, and will encourage scientific integration, coordination of research and the construction and testing of hypotheses. Because Ridge 2000 is a time-limited program, the rapid release of all collected data will benefit all researchers. For more information, the reader is encouraged to read the policy in its entirety on the R2K website:

http://r2k.bio.psu.edu/

On the basis of the MG&G Data Management Workshop recommendations, the Ridge 2000 Program issued an RFP for the August 15, 2000 OCE deadline to create a data management system and data management office. The Ridge 2000 data management system will catalog multiple levels of metadata with user-friendly, web-based tools for searching and accessing data. These include complex searches, relational databases and effective visual display of all types of data. The greatest advances in our understanding of the ocean basins will come from cross-disciplinary investigations that are facilitated by the merging of different data types.

The Integrated Study Site interim websites, the site-specific databases and reference lists are currently under development and represent the first steps towards achieving the free exchange of data. The websites for each of the sites are:

The East Pacific Rise, 8-11°N:

http://www.soest.hawaii.edu/hmrg/Mesotech/EPR_Archive.htm

The East Lau Basin:

ftp://iniki.soest.hawaii.edu/pub/lau/

The Endeavour Segment of the Juan de Fuca Ridge:

http://bromide.ocean.washington.edu/ped/

The Ridge 2000 program recognizes that many of the data management needs of Ridge 2000 and MARGINS are similar and encourages close cooperation between the groups who are developing the respective data management systems. Moreover, Ridge 2000 envisions that the data management systems for the two programs will interface, allowing linked complex searches and perhaps even web-based visualization of linked data. Towards that end a database working group consisting of Ridge 2000 and MARGINS scientists will serve as the oversight and advisory body during the creation of the Ridge 2000 data management system.

A GERM Perspective on MARGINS Databases

Hubert Staudigel[1], and Gray Bebout[2]

1. Scripps Institution of Oceanography, La Jolla, CA 92093-0225, USA; E-mail: hstaudigel@ucsd.edu
2. Department of Earth and Environmental Sciences, Lehigh University, 31 Williams Drive, Bethlehem, Pennsylvania 18015, USA

A GERM Perspective on MARGINS databases.

As MARGINS moves towards an establishment of its databases many important decisions will have to be made on a range of issues, including the structure, philosophy and type of databases and prioritizing funding decisions between different concepts for databases. There will be concerns about the general needs of MARGINS as an integrative scientific initiative and concerns about the needs of particular scientific disciplines. Some disciplines already effortlessly entered the Information Technology (IT) age while others need much help to find ways how their information can be presented and archived in databases. There are important "cultural"differences between disciplines with regard to database use and data publication. Some already have central archives for their data, with a seamless and effective process of data publication and storage while others don't have such archives, and much of their data are never published (e.g. GERM Steering Committee, 2001). Some earth science disciplines are used to interdisciplinary work; others are not. Integration of diverse contributing sciences is crucial to the success of MARGINS and integrated databases are a necessary first step. Careful cost-benefit considerations have to be made. Should the money go to the neediest groups that are furthest away from any IT infrastructure, or should it go to the best organized group with well established IT concepts? When evaluating concepts for databases decisions have to be made on the expected impact on a particular discipline and MARGINS as a whole. What will be the community buy-in? How robust is the planned IT infrastructure and how long-lived is an electronic archive? MARGINS is a microcosm of all of earth sciences and ocean sciences combined, and establishing an effective database infrastructure for MARGINS is not that different from establishing an IT infrastructure for all of the earth sciences. A tall order.

When decisions are made regarding database establishment, it is important for MARGINS to study the database efforts of other organizations in earth sciences to learn from their discoveries, experiences and developments. There is very little use in re-inventing technologies that already work elsewhere or re-paving well beaten paths to failure. The stakes for the planners are high, because there are limited funds and once a database is set up, it will be very difficult to gain the necessary momentum to reorganize its internal structure or metadata content. Learning form other efforts is cost effective, reduces the time of development, and it produces more robust and effective databases for the MARGINS program. Integration is also crucial in this respect since it minimizes duplication in effort, both in the design and implementation of these databases, and in their data population. There are a good number of examples to study. This article describes efforts of the Geochemical Earth Reference Model (GERM) initiative and its current host organization EarthRef.org at Scripps Institution of Oceanography that also operates a series of database services. EarthRef.org, GERM and associated sites offer databases and other IT science service structures that could be emulated by MARGINS or the contributing database developers. This article will also refer to a range of activities where GERM and EarthRef.org had an impact on a range of issues in particular pertaining to data publication and copyright, involvement of the wider community in database activity and metadata approaches.

The Reference Model Approach

The main goal of the GERM initiative is to set up an information technology infrastructure to aid in the development of a state-of-the-art geochemical reference model of the earth. Given the inaccessibility of most of the earth to direct sampling, such a reference model has to be understood more as an optimization process for diverse geophysical and geochemical data towards a reference model, rather than the immediate establishment of a perfectly known reference state of the solid earth. In this process GERM divides the earth up into major geochemical reservoirs, that are mostly defined on basis of geophysics and archives geochemical data on the composition of these reservoirs, and the chemical fluxes between them. The geophysical reference model offers an archival structure that allows GERM to organize very effectively all information that contributes to the understanding of a particular geochemical reservoir (like the earth's core or the lower mantle). This allows reconciliation of geophysical and geochemical data on a reservoir-by-reservoir basis and through mass balances of the whole earth and a comparison with cosmochemical constraints. GERM also supports the development and archival of modeling approaches, like a global mass balances or other models that help in geochemical data interpretation or data reduction. Overall GERM and EarthRef.org aims to provide a complete geochemical working environment, including data archival and data analysis that further our understanding of global geochemistry. It is clearly recognized that geochemical reference models have to be integrated with a geophysical reference models, including low and high temperature geochemistry, cosmochemistry, experimental petrology, seismics, geomagnetism, geodynamics etc.

The situation of GERM is quite similar to the one of MARGINS. MARGINS already defined its “reference models”, in a series of cartoons defining reservoirs, such as the “Subduction Factory” or “Source to Sink”. These reference models are primarily defined based on geophysics data and the geochemistry needs to be integrated into this overall concept. Many reservoirs in MARGINS are hidden and their chemistry and phase mineralogy has to be inferred from geophysics or other constraints. Similarly, many fluxes can be determined only as missing/remaining fluxes from mass balance considerations. The geophysical MARGINS reference models offer scientific guidance for the principles of archival of its data, and a framework for optimizing data. These similarities are not really surprising because MARGINS actually covers a very important part of GERM. Plate margins are the key locales for most relevant earth processes, in terms of fluxes and resources. For this reason it is no accident that MARGINS and GERM held a joint special session at the 2001 Fall AGU meeting.

Community Involvement

GERM and EarthRef.org works with the geochemistry and earth science communities, to guide database development and improvement of the interdisciplinary understanding of large scale earth processes and models. This community involvement includes seeking advice on the state of the art science needs regarding data integration, database structures and metadata formats as well as the direct involvement of the community in database development and data ingestion. GERM holds biannual workshops and special sessions at other meetings to define the key scientific questions in global geochemistry and to develop an IT infrastructure that serves these purposes. The results of these meetings are summarized in a series of workshop reports and publications in the literature (e.g. Staudigel et al., 1999, Bebout et al., 2001, Goldstein and Melson 2001, GERM Steering Committee, 2001; Staudigel et al., 2002; Helly et al., 2002) that can be, along with the abstract volumes, found at:

http://Earthref.org/publications/

Another type of community involvement in database development may be illustrated with the GERM partition coefficient database. R. Nielsen (Oregon State University) developed this database and is responsible for its contents, GERM/EarthRef.org handles the IT infrastructure and the maintenance of the database. GERM also links to databases that support important goals of GERM. Those include in particular the Lamont Ridge database (PetDB; Langmuir, Lehnert, Su) and the Mainz GeoROC database (Hofmann, Sarbas). A persistent goal of GERM and these databases is to establish interoperability amongst these three databases. To this end, GERM worked with these database efforts to establish common metadata and data format standards (e.g. Staudigel et al., 2002).

EarthRef.org also recognized that a database can be maintained in a cost-effective fashion only when the community is actually also involved in identifying high priority contents and in the population of the legacy database with these materials. To this end detailed archival procedures were developed that enable the community to help in the data population of the EarthRef database. These procedures include a detailed data population protocol for the archival of legacy data as well as the nearly automatic uploading of new data publications in well established data/metadata formats.

GERM and EarthRef also realized that it is very worthwhile to work with other earth science disciplines with mutual benefits. The benefit to GERM includes a more effective integration with other disciplines. Collaborating disciplines benefit from using GERM templates for their own database and metadata developments. For example, EarthRef.org was involved to much mutual benefit in the development of archival and display methods for marine geological and geophysical (MGG) data in the SIOExplorer project:

http://SIOexplorer.ucsd.edu

EarthRef.org assisted the paleo- and rock magnetics community in the development of their metadata scheme that can now be seamlessly related to geochemistry metadata (Koppers et al., 2002; Constable et al., 2002; Keizer et al., 2002; Genevey et al., 2002). Some of this work will be presented in a series of database oral and poster sessions at the 2002 Fall AGU meeting that is sponsored jointly by the Geomagnetism/Paleomagnetism and Ocean Science Sections.

The importance of community involvement for MARGINS is quite obvious in its workshop and its science planning activities. To extend this community involvement into the database development is a relatively small step. Integration of database activity with state-of-the-art science and the needs of the community is essential, and this can only be accomplished if database efforts pay substantial attention to the needs of the community.

Metadata

Metadata, data about data, are crucial to all database developments, and almost all database developments naturally converge on discussions about metadata. Metadata carry information about how and where data were obtained or how they can be used. metadata include sample descriptions in terms of their location, relationships to other samples, their type, age, and their relevance to particular scientific themes, etc.. Metadata help characterize data, spell out the rules how data should be handled, and they help to find data. The quality of metadata determines the level of interoperability and utility of databases. It is unthinkable, that a database can be established without detailed attention to metadata. For this reason, metadata definitions are crucial for an interoperable system of MARGINS databases.

GERM established a metadata group and discussed metadata during its 2001 meeting in La Jolla. This effort resulted in two papers of Staudigel et al. 2002 and Helly et al. 2002 (G-cubed in press; see website) that describe a very flexible concept to geochemistry metadata, from a geochemist's and an information technologists' perspective, respectively. Geochemical metadata were structured in a hierarchical organization that include blocks of metadata that focus on particular themes, such as location, sample description, or analytical techniques. The metadata definitions also include a set of text/keyword metadata that relate the data to particular reservoirs or fluxes in the GERM earth reference model and other keywords that allow for linkage of data to particular portion of the model, to tock types, tectonic settings etc (Staudigel et al. in press). These types of metadata play an important role in the linkage of diverse data types to a particular region or a particular part of a reference model.

The modular nature of metadata blocks of Staudigel et al. (2002) and Helly et al. (2002) allows easy transfer of metadata formats to other disciplines (like sample location or description) while others remain specific for a particular sample type or discipline. EarthRef.org was able to adapt the geochemistry metadata template to the needs of paleo- and rockmagnetics community. These metadata templates were first presented in a May 2002 workshop in La Jolla, and refined in another workshop in Minneapolis in September 2002 (see EOS article to be printed before Fall AGU meeting). The near-final metadata set is fully interoperable with the geochemistry metadata, and it is likely to be finalized at an international paleo- and rockmagnetic's database workshop to be held in April 2003 in Nice, France. Similar efforts were made by the Marine geology and geophysics (MGG) group Geological Data Center at the Scripps Institution of Oceanography, which already are heavily used in the database structure for MGG data in the NSDL sponsored SIOexplorer project. The metadata design for this project is partially discussed in the Helly et al. (2002) paper in G-cubed.

Staudigel et al (2002) also made recommendations for which metadata should be "mandatory", the minimum amount of metadata necessary for a scholarly publication of data. Such a decision is important, because consistent use of metadata is important for effective data searches or data operations within a database. After some reflection, the decision of what metadata one considers "essential" is typically quite obvious, but such decisions have to be made, finalized and embraced by the science community, and enforced by the editorial and science funding process.

Helly et al. (2002) focused on interoperability of data bases and devised a metadata interchange format for this purpose. A Metadata Interchange Format (designated by the extension ".mif", read "dot mif") and a metadata transport format (.mtf) describe the structure and the metadata contents of any computer file and allow for their analysis and exchange with other databases. Such computer files can be any type of digital object (e.g. maps, movies, MP3's, seismic raw data, geochemical data tables, or data bases themselves) and, therefore, are designated Arbitrary Digital Objects (ADOs). ADOs may comprise entire database tables for easy and fast exchange of both data and metadata between multiple databases. These MIF files can also be readily transferred into XML. Effective transfer of metadata is a pre-condition for any interoperability between databases connecting any type of ADO.

The MARGINS initiative critically depends on integration of very diverse metadata and data types. However, metadata developments in EarthRef.org could relatively simply be taken over by MARGINS, or at least it could serve as a template for development of a specific set of metadata. The only rigid principles are the need for a scholarly archival, and the need for interoperability of databases. The grouping of metadata into sets and their hierarchical structure allows for the relatively easy development of new metadata formats for other disciplines, and for the translation between alternate metadata concepts.

However, it has to be emphasized that the database integration needs of MARGINS are unprecedented in earth sciences, whereby on-land information has to be integrated with off shore data, and very diverse disciplines such as low and high temperature geochemistry has to be integrated geophysical and geological data from exploration and monitoring studies.

What should be archived?

There is a wide range of data and metadata in earth sciences that should be archived, legacy data and new data, data printed in journal articles or books and data that are distributed electronically. Archival may include data and derived data or other data products. The principles of scholarly science conduct requires archival of all components of a paper that is necessary to reproduce, verify or falsify any conclusion. This means archival of the complete chain of products that defines a scholarly science trail, including data, derived data and the derivations or operations that lead to a particular scientific product that will be archived in the public record of peer reviewed science publications. All critical links in this chain should be archived. The actual use of the archived products will vary. In MARGINS, it is quite likely that derived data products are more useful to the broad community than the underlying data. Such data products include bathymetric maps, the locations of earthquakes and their fault plane solutions, or a seismic section. The original data , however, have to be archived as well because they are the primary source of all derived products.

GERM and EarthRef.org established a database structure that allows for archival of all kinds of data, ranging from individual data points to synthesis data, and including any other arbitrary digital object (ADO), including, diagrams, images, maps, videoclips etc.. EarthRef.org is working with the San Diego Supercomputing Center (SDSC) that offers a range of archival options for small and very large ADOs in the EarthRef Digital Archive:

http://Earthref.org/databases/ERDA/

GERM includes some very simple databases, such as the "reservoir database" that is probably best described as a simple matrix of element number vs. reservoir, containing concentration data and a literature reference. The latter literature reference links the user to the EarthRef "reference database" where each paper is archived to reflect the original source of each data point in this database. This is done by supplying scans (in the JPG and PDF format) and their digitized forms (in the Excel or Word format) for all abstracts, data tables and technical notes. These can be viewed and subsequently downloaded from the EarthRef.org website. All in all, this assures complete traceability of data from its source to each record in the database. A more complex structure underlies the "seamount catalog" at EarthRef.org that is designed to archive a wide range of data, including maps, marine geophysical data, and any other data type pertaining to a particular seamount. EarthRef.org was also involved in the development of the "SIOexplorer" and the "Bridge the Gap" initiative between SIO and the UCSD libraries, where up-to-date oceanographic information is displayed in connection with SIO library contents, such as its archives of ancient shiplogs or photographs.

GERM also recognized that it is essential to allow for a seamless connection between data and the modeling of data. For this reason, GERM began to archive geochemical modeling tools at:

http://Earthref.org/tools/

These tools may either be developed at SIO (ArArCALC; Koppers, 2002) or elsewhere (e.g. energy conserving R-AFC modeling; e.g. Spera & Bohrson, 2002) and are supplied with additional information such as online help tools, installation procedures and example files.

The diverse contents to be archived in MARGINS databases requires the use of many different types of archival structures, quite analogous to GERM and the effective use of MARGINS data would make it very desirable to archive modeling approaches as well as data. Open and easy access to data and models is critical to the success of MARGINS.

Data Publication Issues

Database development and data publication are closely linked. Very few earth science disciplines use well organized data repositories for publication of their data (e.g. IRIS, or NGDC). Some disciplines rely on an informal archival of data amongst major institutions or on servers associated with major research groups. A large number of earth science disciplines, however, has no central or distributed data archival facility and relies exclusively on the publication of data in the peer reviewed literature. This is true in particular for non-data intensive disciplines like geochemistry, geology, petrology, volcanology, but also there are no central data archives for many data types in seismology (multichannel, seismic monitoring), non-marine gravity or magnetics or electromagnetic surveys. All these disciplines fare rather prominently in MARGINS.

Publication of data in the peer reviewed, printed literature has many intrinsic problems that can interfere substantially with the goals of science and in particular with the value and efficiency of database activities:

1. Paper (or its electronic equivalent) limits data publications in several ways :

   - Many high profile letter journals encourage very short articles and actuall discourage data publications (e.g. Nature, Science)
   - Space limitations and aesthetics of page layout limit the publication of data, and eliminate the publication of metad ata entirely
   - Often, data are published in graphical form only
   - Publication of data on paper involves much unnecessary work, first formatting it for human eye consumption in page-layout, and then bringing it back to a computer legible form either by re-typing of optical character recognition.

GERM has been instrumental in resolving many of these issues, by encouraging the use of digital data/metadata supplements and the encouragement of editorial practices that force the publication of all data contributing to a paper (GERM Steering Committee, 2001). MARGINS has now the opportunity to establish publication policies and/or data archival resources that encourage the scholarly archival of data published in all MARGINS papers, in formats that allow for easy re-use of the data by the science community.

2. A large body of high quality data is never published because they don't lead to the type of papers that are publishable in first-rate journals. Such data include results that appear uninterpretable in our current state of knowledge or data reduction techniques, data that merely confirm existing hypotheses, or data sets there are not large enough to warrant an article in a high profile journal. Such data provide a valuable resource to science, because they might eventually lead to new discovery, in context of new, additional evidence or after a change in paradigm, or the arrival of new modeling techniques. Such data have to be archived and made available to science. The problem was addressed during the development of G-cubed by a group of GERM scientists: G-cubed has a speicific publication type for such data, "data briefs". This paper category is specifically designed for publication of data, without a major requirement for an in depth discussion. This type of publication will be a useful mechanism to

archive such materials produced from MARGINS research projects.

3. Many earth science data are typically published in inconsistent formats and critical metadata are inaccessible, either by not being reported at all or available only by extraction from text passages. The GERM Steering Committee (2001) encouraged the use of electronic supplements, and the adoption of consistent data/metadata formats that would allow for direct uploading into databases. Major changes in publication habits are needed and MARGINS could work towards such a goal as well.

4. The high costs of scientific publications and copyright problems limit the distribution of science and the use of papers on websites for most first-rate scientific journals. To address this problem, GERM scientists and the AGU formulated a copyright policy for G-cubed that gives community databases like GERM the right of unlimited use and free display of GERM-related publications from G-cubed. A MARGINS database could fall in the same category as GERM.

GERM and the MARGINS Initiative on Convergent-Margin Dynamics

GERM and MARGINS are currently working on a number of common projects, integrating workshops, database activity and publication efforts relating in particular to the development of the Subduction Factory Initiative and its more focussed pursuits. The GERM subduction pages were optimized for synergy with the SubFac initative, resulting in part from significant overlap among the individuals contributing to the planning of the two entities. The Subduction Factory focus sites are among the eight GERM-subduction flux focus arc-trench systems and GERM will prioritize efforts aimed at the IBM and Central American convergent margins. As noted above, special sessions at Goldschmidt, European Union of Geosciences, and AGU meetings have been co-sponsored by GERM and MARGINS participants, most recently in a GERM and MARGINS/Subduction Factory session at the Fall, 2001, AGU meeting in San Francisco ("Trench to Subarc: Diagenetic and Metamorphic Mass Flux at Convergent Margins"; see abstract volume at:

http://earthref.org/publications/main.htm

A theme in G-cubed, devoted to this special session:

http://g-cubed.org/theme/theme.shtml

will exploit the close interfacing possible between GERM and MARGINS. All papers of this theme will be available freely on GERM/EarthRef. A number of the scientists who participated in ODP Leg 185 (drilling in the IBM focus site aimed at understanding chemical flux in subduction zones) were actively involved in Subduction Factory and GERM planning, and a separate G-cubed "theme" aimed at chemical flux in the IBM system ('G-cubed theme "Oceanic Inputs into the Subduction Factory"'. The plan, based on a number of discussions at GERM planning workshops (and at other meetings, including SubFac planning meetings), has been for the GERM subduction pages to sponsor the development of a conceptual (and semi-quantitative?) model for chemical mass-balance across individual arc-trench systems, focussing first on the two SubFac focus sites (see the "straw man" model presented on the present GERM subduction web pages). Overall, the existing database structure of EarthRef presents many exciting opportunities for presentation and creative combination of interdisciplinary observations to understanding of mass and energy transfer at convergent margins [i.e., not just geochemical data]. EarthRef/GERM and MARGINS/Subduction Factory have many of the same goals and would mutually benefit from as close an interaction as possible.

Final Thoughts

The needs and requirements for databases in earth sciences and MARGINS are extremely diverse and any successful database effort should maximally allow for an effective cross-communication between disciplines. Databases should carry data and data products, serving the specialists and the broad community. Data archival and data publication should not be separated. Attention should be paid to metadata, and for MARGINS it would be beneficial to have a data / metadata organization that focuses on some MARGINS reference models such as the Subduction factory or Source-to-Sink. It is quite obvious that the MARGINS database needs are not Rocket Science, most of it can be done with existing and well established IT technologies. With very few exceptions, data volume and data transfer rates are not a challenge, and our capabilities are rapidly expanding. The key to success of GERM, and to MARGINS is an effective community buy-in rather than serving a small specialist community and widest possible integration with the larger MARGINS science community. Scientific guidance by the broadest and highest level science community possible has to be the main governing principle of database structure. There has to be a very intimate involvement between science community and database development and database activity. This is not done by a user group whose members may familiarize themselves with the current database issues on the flight to the annual meeting. True community buy-in is hard-earned but necessary.

Acknowledgements

This material is based on collaborations with colleages at UCSD (A.A.P Koppers, J. Helly, S. Miller, C. Constable, Neal Driscoll , C. Johnson and L. Tauxe), members of the GERM Steering Committee of (D. Anderson, L. Derry, W. McDonough, A. Hofmann, W. White, H. Shaw A. Zindler), the other co founders of G-cubed (C. Langmuir, W. McDonough, H. Shaw, W. White, A. Zindler), and some of the developers of database efforts at Lamont (K. Lehnert) and at Mainz (B. Sarbas).

References

Bebout, G., J. Blichert-Toft, R. Carlson, K. Cooper, M. Evans, K. Kelley, J. Konter, C. Reif, O. Rouxel, A. Shaw, G. Shields, H. Staudigel, and D. Stegman, 2001, Report on GERM 3, La Jolla, March 6-9, Newsletter of The Geochemical Society, Vol 108, p 17-19 (http://gs.wustl.edu/archives/gn/gn108.pdf)

Constable, C. H Staudigel, L Tauxe, A Koppers, C Johnson, P Solheid, M Jackson, S Banerjee, S Pisarevsky 2002, New Magnetic Database Initiatives: Exploitation of and Integration with other Developments. , EOS Fall Meeting of the American Union of Geophysics.

GERM Steering Committee, 2001. Electronic data publication in geochemistry: A plea for "full disclosure," Geochem. Geophys. Geosyst., vol. 2, Paper number 2001GC000234 [1026 words]. Published October 11, 2001.

Goldstein, S.L., W. Melson, 2001, Geochemistry in the 21st Century: a new GERM initiative, Newsletter of The Geochemical Society, Vol 108, p 19 -21 (http://gs.wustl.edu/archives/gn/gn108.pdf)

Helly J., H. Staudigel, and AAP Koppers, in press, Scalable Models of Data Sharing, G-cubed, http://earthref.org/metadata/GERM/main.htm

Helly, J J , H Staudigel, A Koppers, 2002, Scalable Models of Data Sharing in the Earth Sciences., EOS Fall Meeting of the American Union of Geophysics.

Keizer, P., A Koppers, L Tauxe, C Constable, A Genevey, H Staudigel, J Helly 2002, PMAG: Relational Database Definition., EOS Fall Meeting of the American Union of Geophysics.

Koppers, A , C Constable, L Tauxe, H Staudigel, J Helly 2002, PMAG: Database Development Under the EarthRef.org Umbrella Website., EOS Fall Meeting of the American Union of Geophysics.

Staudigel H., Albarede, F., J. Blicher-Toft, J. Edmond, B. McDonough, S.B. Jacobsen, R. Keeling, C.H. Langmuir, R.L. Nielsen, T. Plank, R. Rudnik, H.F. Shaw, S. Shirey, J. Veizer, and W. White., 1998, Geochemical Earth Reference Model (GERM): Description of the Initiative. Chemical Geology, 145: p 153-161.

Staudigel H., B. McDonough, Albarede, F., G. Maters, H.F. Shaw, A. Lasaga and W. White, 1998, GERM Workshop La Jolla, California March 10-13, 1998, Abstract Volume and Current State of the GERM Model, 181pp.

Staudigel H, B. McDonough, H.F. Shaw, 1998, Second GERM Workshop held in La Jolla, Californai, March 11-13, 1998, The Geochemical News 96: 22-23.

Staudigel, H., J. Helly, A.A.P. Koppers, H. Shaw, B, McDonough, A. Hofmann, C. Langmuir, K. Lehnert B. Sarbas, L. Derry, and A. Zindler, in press, Electronic data publication in Geochemistry, technical note G-cubed, http://earthref.org/metadata/GERM/main.htm

H Staudigel, A A Koppers, C Constable, J Helly 2002, GERM in EarthRef.org: A reference model approach to data bases., EOS Fall Meeting of the American Union of Geophysics. Energy-Constrained Open-System Magmatic Processes III:

Spera, F. and W.A. Bohrson, Energy-Constrained Recharge, Assimilation and Fractional Crystallization (EC-RAFC) 2002 in press, technical note G-cubed.

Hubert Staudigel *is Chairman of the GERM Steering Committee*
Gray Bebout *is GERM Subduction Zone co-editor and MARGINS liaison*

MARGINS Newsletter No. 8, Spring 2002

MARGINS Mailing List Update

Cut or copy this form and mail it to: MARGINS Office, Lamont-Doherty Earth Observatory, P.O. Box 1000, Palisades, NY 10964, USA, or make changes online at http://www.ldeo.columbia.edu/margins

Name:

Department:

Organization:

Street Address:

City: State: Country:

Postal Code: Telephone:

E-mail: Fax:

Your Home Page URL:

Your Department's Web Site URL:

Towards a Data Management Infrastructure for MARGINS: Examples from Boomerang 8 and the Virtual Research Vessel

Dawn J. Wright

Department of Geosciences, Oregon State University, Corvallis, OR 97331-5506, USA; E-mail: dawn@dusk.geo.orst.edu

Introduction

Large interdisciplinary Earth science programs such as MARGINS, RIDGE 2000 (R2K), GLOBEC, NASA's Earth Observing System, and others, are currently focusing not only new scientific insights via collaborative research and multidisciplinary studies, but on a more standardized exchange of data between individual projects, styles of data presentation and analysis, and the quality of supporting metadata. One key to developing and maintaining unified community databases for these programs lies in building and supporting a general organizational structure linking distributed databases through the web. Described here are two end-member examples of data access and management. One is a simple web site, merely linking the user to data sets for download. The other is a complex "computational environment" employing three technologies: web GIS, a computational experiment management system, and a relational database management system (RDBMS). Both projects may be of interest to MARGINS researchers and data managers.

Boomerang 8

In 1996 the Oregon State University Department of Geosciences, in collaboration with the Scripps Institution of Oceanography, developed the Boomerang 8 online data archive at sea aboard the R/V Melville. Boomerang Leg 8 was an ODP site survey (Bloomer and Wright, 1996), and addressed a number of scientific issues concerning the origin and structure of the Tonga forearc and trench including: (1) testing the hypothesis that the forearc is comprised of an ophiolitic basement, formed in the earliest stages of subduction by high-volume, short-lived arc volcanism (Bloomer et al., 1996; Kelman, 1998); and (2) constraining the mechanisms of tectonic (subduction) erosion along the trench (e.g., Hussong and Uyeda, 1981; Hilde, 1983; Bloomer and Fisher, 1987; Wright et al., 2000), and how its effects may be distinguished from the subduction of the Louisville Ridge. As a site survey, Boomerang Leg 8 was tasked with exploratory bathymetric, sidescan, and single-channel seismic reflection surveys over proposed ODP drill sites, as well as a comprehensive dredging program. The resulting web site is very simple by today's "standards" (i.e., no advanced features such as expandable, collapsible pull-down menu interfaces, web frames, or online mapping are included), but was unusual for its time in that it was designed and coded entirely at sea, and then tarred into an archive for easy transfer to a permanent, shorebased server (Figure 1):

http://buccaneer.geo.orst.edu/dawn/tonga/

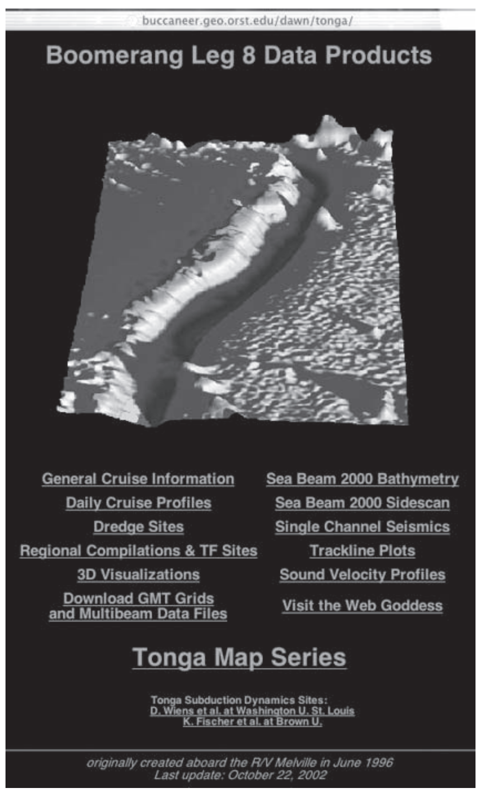

*Figure 1. Screen shot of the web portal to data products from Boomerang 8 for the Tonga forearc and trench (*http://buccaneer.geo.orst.edu/dawn/tonga/*).*

where it has since been used by a small community of researchers. The site contains scores bathymetric and sidescan maps, multibeam data files and grids, postscript files of selected single-channel seismic lines, dredge sampling maps, bathymetric profiles, trackline plots, sound velocity profiles, and 3-D visualizations. In a similar vein, data from the Eastern Lau Spreading Center are available via an anonymous ftp site in development at the University of Hawaii:

ftp://iniki.soest.hawaii.edu/pub/lau

The Virtual Research Vessel

The Virtual Research Vessel (VRV) is a research collaborative of the University of Oregon, Oregon State University, and the Evergreen State College that is much more ambitious than the Boomerang 8 archive. VRV incorporates a web-based geographic information system (GIS) for viewing, loading, and selecting subsets of data and metadata, but also a separate relational database management system (RDBMS) and application programming interfaces (APIs) to support the coupling of numerical models. The project is fully described in Wright et al., (in press) and may be visited on the web at:

http://oregonstate.edu/dept/vrv/

Rather than a web site for simple access and download, VRV is a “computational environment” with the primary goal of allowing researchers to computationally link disparate data, as well as numerical simulations, so that they may actually undertake interdisciplinary experiments online. Researchers thus have the ability to explore new relations between observables collaboratively over the web, quantitatively evaluate hypotheses, and refine numerical simulations. An additional goal is the ability for researchers to build self-consistent models of complex phenomena from existing models of isolated phenomena. The prototype, a work in progress, provides a case study for the R2K community as to whether this coupling of data, maps, and models will work toward the desired goal, and includes test databases primarily from the East Pacific Rise at 9-10°N.

The web GIS in VRV is based on ESRI’s ArcIMS (Internet Map Server)®, and features visual displays with the ability to incorporate diverse data sets at different resolutions and scales and to create dynamic maps (Figure 2). For example, it is possible to locate hydrothermal vent sites on maps, query or contour for variables such as water temperature or rock composition, to interact with tabular and grid-based data sets, and to simultaneously download both data and metadata for input to desktop software (e.g., ArcView, Matlab). Here logical queries can be made and spatial relationships can be seen between various layers or themes of data. This provides a rapid, “unselfish,” and logical way of disseminating knowledge for rapid response to such mid-ocean ridge events as mega-plumes or volcanic eruptions indicated by seismic events. In addition, images or video clips may be incorporated via a “hotlink” to locations on maps. And there exists the potential for simultaneous access to both data from a web server and local data from the client’s desktop, along with the ability to dynamically edit and annotate maps.

Simple analytical capabilities are available as well. Users may create or

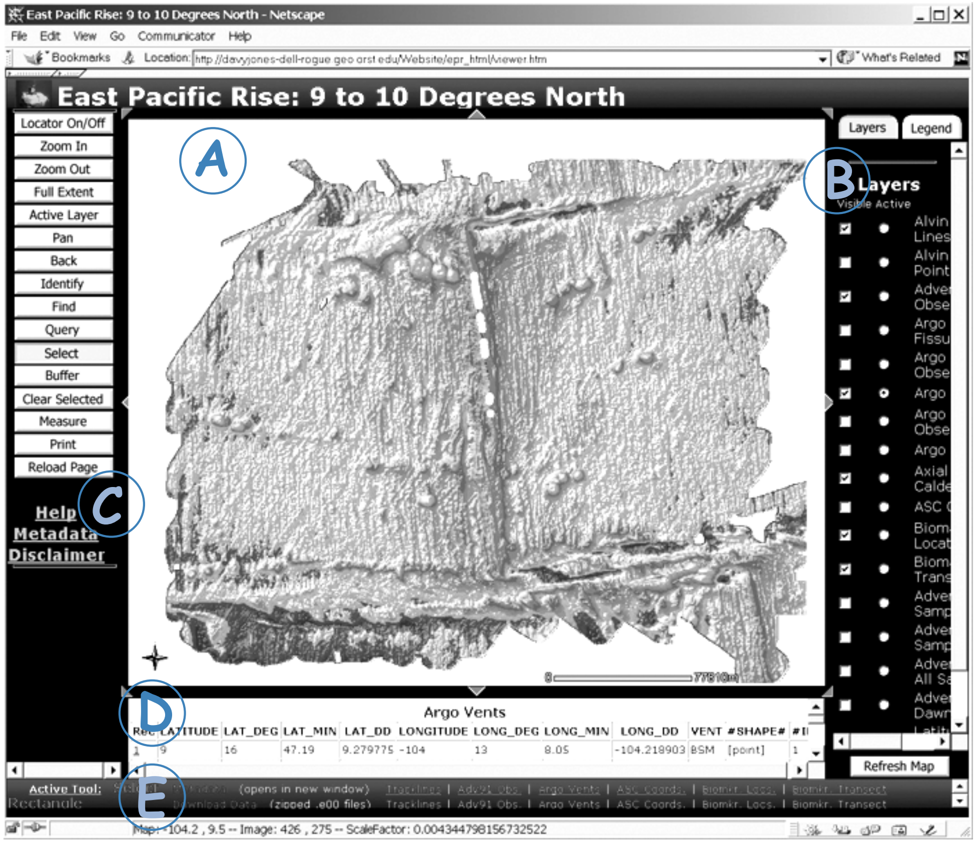

*Figure 2. Main interface of the web GIS for the Virtual Research Vessel (VRV) prototype (*http://oregonstate.edu/ dept/vrv/tools.html#gis*). Features of the interface include (A) creation and access of maps at varying scale where other data can be overlain, such as the active map layer of vent locations in yellow (B). (C) help for users, as well as all metadata open in separate windows. (D) all mapped variables are tied to a GIS database where all attributes are immediately available and can be searched and queried. Right below (D) is a bar where all metadata and data can be downloaded directly to the user’s desktop in zip files. (E) as the user moves the cursor around the map, the lat/long pixel position, and scale factor of the map are displayed.*

access maps of different resolution at different scales as demonstrated in Figure 3 where the user has zoomed into large detail to map the axial summit trough at 9N with vent locations (two of which have been buffered out to a distance of 50 m) and Alvin submersible tracklines. The user may also measure distances in meters between selected points as shown by the red line measuring the distance between two critical vent sites in Figure 3. There is also interaction with tabular data to show the attributes of points, lines, and areas mapped as well as metadata for all layers that can be mapped. What can easily be added to this existing interface is the ability to contour maps at various scales for variables such as water temperature or rock composition. All data are available through this web interface as downloadable zip files.

Work in progress on the VRV prototype includes efforts to develop: 1) a virtual database to incorporate diverse data types (along with domain-specific metadata) into a global schema, allowing for web-query across different marine geology data sets, and an analogous, declarative (database-available) description of tools and models; 2) the ability to move data between the GIS and the RDBMS, along with the tools to encourage data submission to archives; 3) tools for finding and viewing archives, and translating between formats; 4) support for “computational steering” (tool

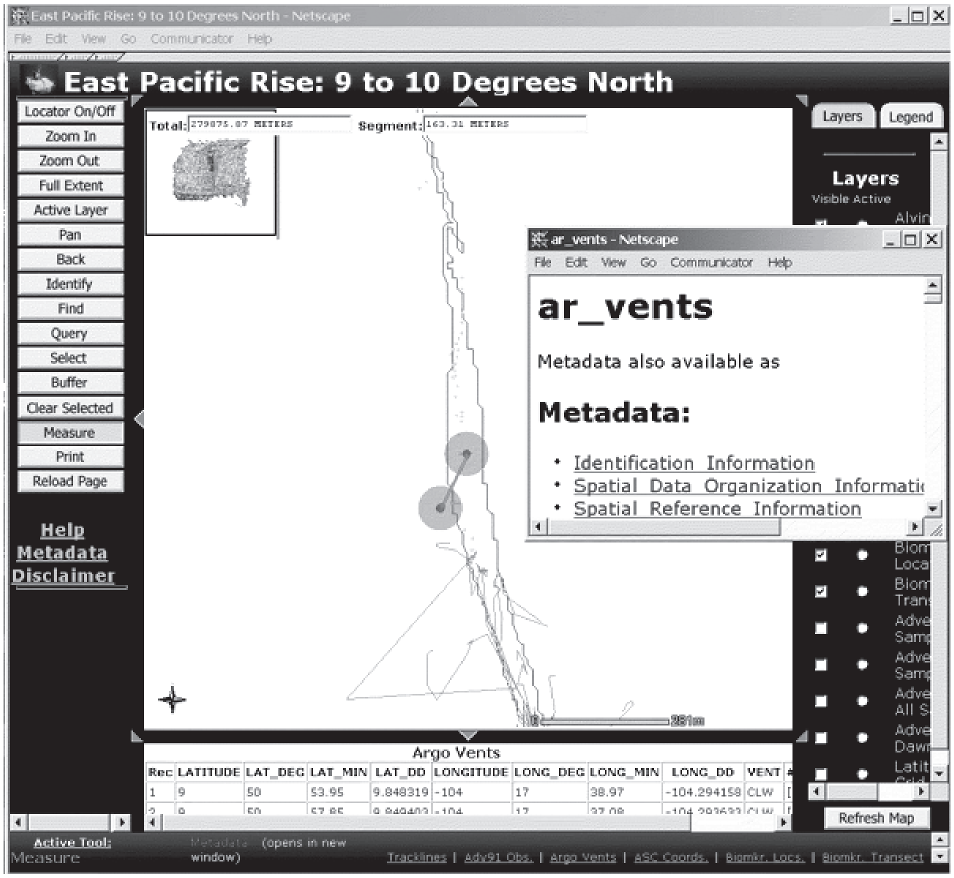

Figure 3. Simple analytical capabilities (e.g., distance measurement, buffer generation) of the prototype web GIS for the VRV prototype as described in the text.

composition) and model coupling (e.g., the ability to run tool composition locally, but to access input data from the web), APIs to support the coupling, especially of programs that are running, and the writing of data wrappers to publish programs; 5) support of migration paths for prototyped model coupling; and 6) export of marine geological data and data analysis to the undergraduate classroom (VRV-ET, "Educational Tool").

Conclusion

It is hoped that the examples above will be useful to the MARGINS community in their continued development of data management infrastructure for MARGINS Focus Sites, an infrastructure that includes web portals to key data sets and metadata, maintained and updated by professionals familiar with the various data sets, and containing sufficient information to be broadly useful across all the disciplines encompassed by MARGINS scientists.

And there are several other web portals in existence, providing mid-ocean ridge data to R2K researchers, that may be also be of interest to the MARGINS community. These include results from the AHA-NEMO cruises at Woods Hole Oceanographic Institution:

http://science.whoi.edu/ahanemo2

the multibeam and petrologic data sets at Lamont-Doherty Earth Observatory:

http://ocean-ridge.ldeo.columbia.edu
http://petdb.ldeo.columbia.edu

the acoustic and photographic online archive available at University of Hawaii:

http://www.soest.hawaii.edu/HMRG/EPR/index.htm

the Scripps Institution of Oceanography's Ocean Exploration data portal:

http://sioexplorer.ucsd.edu

and the University of Washington's Endeavour GIS and Portal to Endeavour Data (PED) sites:

http://bromide.ocean.washington.edu/gis

http://bromide.ocean.washington.edu/ped

Acknowledgments: The author gratefully acknowledges the support of NSF grants OCE-9521023 and OCE-0081487, the support of Todd Porteous and Stu Smith at Scripps in the building of the Boomerang 8 data archive, and the ongoing labors and collaborative support of Virtual Research Vessel colleagues Jan Cuny and Douglas Toomey of the University of Oregon, and Judith Cushing of Evergreen, as well as Margo Edwards of the Hawaii Mapping Research Group, and Dan Fornari of WHOI.

References

Bloomer, S. H., and Fisher R. L., Petrology and geochemistry of igneous rocks from the Tonga Trench: Implications for its Structure. J. Geology 95, 469-495, 1987.

Bloomer, S. H., and Wright, D. J., Summary of Site Survey Cruise Results, Boomerang Leg 08, in Support of Proposal 451, Ocean Drilling in the Tonga Forearc: Subduction Geodynamics, Arc Evolution, and Deformation Processes at a Non-Accretionary Convergent Margin, Ocean Drilling Program Site Survey Data Bank, Palisades, New York, 1996.

Bloomer, S. H., Wright, D. J., and Boomerang Leg 8 Shipboard Scientific Party, Geology of the Tonga Forearc: A supra-subduction zone ophiolite, Eos Trans. AGU 77, F325, 1996.

Hilde, T. W. C., Sediment subduction versus accretion around the Pacific, Tectonophysics 99, 381-397, 1983.

MARGINS Steering Committee Highlights

Olaf M. Svenningsen

Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, New York 10964, USA; E-mail: olafs@ldeo.columbia.edu

The MARGINS Steering Committee (MSC) convened at the New Otani Hotel in Honolulu, Hawaii on September 13th and 14th, 2002 in connection with the IBM 2002 workshop. The agenda included the following items:

1. Opening the meeting, introduction of the new Steering Committee members: Scott Linneman, Rudy Slingerland, and Patricia Wiberg.
2. MARGINS funding: present and future funding levels.
3. NSF Program Manager's report on the state of MARGINS.
4. Workshop reports (see this newsletter p. 7 and 12).
5. Upcoming MARGINS events (see announcements in this newsletter).
6. MARGINS Focus Site database generation (see articles in this newsletter).
7. MARGINS presentation materials.
8. Science Plan reviews.
9. EarthScope.
10. New MARGINS Chair: Julie Morris (see Chairman's note on page 4).
11. Steering Committee rotations (see Chairman's note on page 4).

Many of these items are covered elsewhere in this newsletter. Selected items that are not covered elsewhere are summarized below:

1. This was the first meeting for three members of the Steering Committee: *Scott Linneman* of Western Washington University will be the Education and Outreach representative on the MSC, and *Rudy Slingerland* of Penn State University, and *Patricia Wiberg* (who was absent) of Virginia University both represent the Source-to-Sink Focus Initiative.

2. As described in the Spring issue of the MARGINS Newsletter (#8), the funding level for MARGINS was lower than expected. The MSC expressed that contribution from the three NSF divisions sponsoring MARGINS (EAR, OCE, and ODP) should become more balanced. Presently, roughly half of the PI's in MARGINS are EAR scientists, but the EAR division currently contributes less than one fifth of the funding for MARGINS. A letter from Dr. M. S. Leinen, Assistant Director of the Directorate of Geosciences (which includes OCE and EAR), acknowledged the important, if not unique interdisciplinary aspects of the MARGINS Program and reemphasized as a high priority the desire that NSF-EAR should be an equal fiscal partner within MARGINS.

3. NSF is aiming to make awards larger and for a longer duration. The present average award is $100,000 for 2.3 years. NSF also recommends PI's to contact their Program Officer for consultation before submitting collaborative proposals. NSF is also working to make the proposal review process completely electronic.

   NSF also expressed a wish to obtain "milestones" or "nuggets" (short text snippets, preferably accompanied by illustrations that briefly summarize important progress or breakthroughs within each project or program) as an aid in the evaluation of the various components of the MARGINS Program. This discussion continued seamlessly into Item 7:

7. The MARGINS Office was commissioned to produce presentation brochures for the four Focus Initiatives — Rupturing Continental Lithosphere, Seismogenic Zone, Source-to-Sink, and the Subduction Factory — and one general version covering the entire MARGINS Program. At the time of writing this, the MARGINS Office is collecting material for, and designing these brochures, that will also be available as PDF files on the web site.

8. It was reiterated by the NSF representatives that the science plans for the Focus Initiatives should be published as soon as possible. The publication of the science plans doesn't "lock" them, and they will continue to be living, changing documents. The concerns about the competitiveness of the Source-to-Sink science plan were discussed and the need for a rewrite of the document was emphasized. The revised science plan was released for community comments on October 4th, 2002 (see the Source-to-Sink pages at the MARGINS web site).

9. The MSC intends to issue a statement to EarthScope about the integration of efforts, offshore capabilities and more. MARGINS endorses EarthScope and its goals completely. This is particularly important for the deployment of OBS's for the USArray experiment when the "footprint" extends into offshore regions. At present, requests for a significant number of instruments cannot be met and thus it is crucial at this stage to recognize the problem and secure funds for the augmentation of the OBS facility.

MARGINS Town Meeting: The Source-to-Sink Science Plan Revisited at the AGU Fall Meeting 2002

The MARGINS Office is organizing an AGU Town Meeting to discuss the recently revised Source-to-Sink science plan and the fiscal realities of the MARGINS program with the S2S community, cognizant NSF program directors and members of the MARGINS Steering Committee.

This important Town Meeting will be held at 7.00 PM on Saturday, 7th December, 2002, in Room 104 of the Moscone Convention Center. The time and location of the Town Meeting is also posted on multiple pages at the MARGINS web site. Refreshments will be served.

InterMARGINS seeks input for its recently revamped web site

www.intermargins.org

The InterMARGINS Office is always grateful to receive news of upcoming open meetings and workshops relevant to margins research worldwide.

Equally, the Office wishes to post information on its site about all recent and scheduled margins-related cruises wherever they may take place. The Office can be mailed at:

intermargins@soc.soton.ac.uk

MARGINS-NSF logo use reminder

Any researcher presenting results generated by funding from the NSF-MARGINS Program is required to acknowledge this by displaying the MARGINS-NSF logo in the presentation and/or poster. The logo can be downloaded in several different file formats and resolutions from the MARGINS web site:

http://www.ldeo.columbia.edu/margins/Logos.html

(Subduction Dynamics Workshop Report, continued from page 14):

medium with viscous, elasto-plastic and brittle deformation mechanisms. An important class of current models by-passes these concerns by imposing the slab kinematics, but even with this simplified condition many difficulties remain to be solved, in particular regarding the influence of fluid flow on the dynamical properties of the wedge. Several people expressed the hope that some form of comparison between kinematic and dynamic models would be included in an eventual benchmark.

The late afternoon poster session provided another forum for the informal exchange of ideas. One observation from the poster session was that while the new temperature-dependent viscosity, thermal wedge models are similar, there are some interesting differences between models. Several of the posters presented additional details related to the computational methods. Abstracts of the posters are available at the workshop's website.

On Sunday morning, the group returned to begin the task of formalizing a set of benchmarks. The summary of the discussion was formatted into a proposed benchmark by Peter van Keken and has been uploaded onto the workshop web site. Interested persons can view the benchmark proposal and participate in the benchmark, following the instructions on the website. As results become available, they will be posted on the workshop website.

(References for Wright, continued from page 36):

Hussong, D. and Uyeda S., Tectonic Processes and the History of the Mariana Arc: A Synthesis of the Results of Deep Sea Drilling Project Leg 60, in Hussong, D. M., Uyeda, S., et al. (eds.) Initial Reports of the Deep Sea Drilling Project, 31, U.S. Govt. Printing Office, Washington D.C., 909-929, 1981.

Kelman, M., Hydrothermal Alteration of a Supra-Subduction Zone Ophiolite Analog, Tonga, Southwest Pacific, Master's Thesis, Oregon State University, Corvallis, Oregon, 1998.

Wright, D.J., Bloomer, S.H., MacLeod, C.J., Taylor, B., and Goodliffe, A.M., Bathymetry of the Tonga Trench and forearc: A map series, Marine Geophysical Researches, 21(5): 489-512, 2000.

Wright, D.J., O'Dea, E., Cushing, J.B., Cuny, J.E., and Toomey, D.R., Why Web GIS may not be enough: A case study with the Virtual Research Vessel, Marine Geodesy, 26(1-2), in press, 2002.

MARGINS Steering Committee Members

Garry Karner, Chairman
Lamont-Doherty Earth Observatory
61 Route 9W, P.O. Box 1000
Palisades, New York 10964
Tel: (845) 365-8355
e-mail: **garry@ldeo.columbia.edu**

Julie Morris, Chair-elect
Department of Earth and Planetary Sciences
Washington University in St. Louis
Campus Box 1169
One Brookings Drive
Saint Louis, MO 63130
Tel: (314) 935-6926
e-mail: **jmorris@levee.wustl.edu**

Geoffrey Abers
Department of Earth Sciences
Boston University
5 Commonwealth Avenue
Boston, MA 02215
Tel: (617) 353-2616
e-mail: **abers@bu.edu**

Frederick M. Chester
Department of Geology & Geophysics
Mail Stop 3115
College Station, TX 77843
Tel: (979) 845-3296
e-mail: **chesterf@geo.tamu.edu**

Rebecca J. Dorsey
Department of Geological Sciences
1272 University of Oregon
Eugene, OR 97403-1272
Tel: (541) 346-4431
e-mail: **rdorsey@darkwing.uoregon.edu**

Marc M. Hirschmann
Department of Geology and Geophysics
108 Pillsbury Hall
University of Minnesota
Minneapolis, Minnesota 55455
Tel: (612) 625-6698
e-mail: **Marc.M.Hirschmann-1@umn.edu**

Greg Hirth
Department of Geology and Geophysics
Woods Hole Oceanographic Institution
Woods Hole, MA 02543
Tel: (508) 289-2776
e-mail: **ghirth@whoi.edu**

Scott Linneman
Geology Department and Science Education Group
Western Washington University
516 High Street
Bellingham,WA 98225-9080
Tel.: (360) 650-7207
e-mail: **scott.linneman@wwu.edu**

Anne Meltzer
EarthScope liaison
Dept. of Earth & Environmental Sciences
Lehigh University
31 Williams Drive
Bethlehem, PA 18015-3188
Tel.: (610) 758-3660
e-mail: **asm3@lehigh.edu**

John Milliman
Virginia Institute of Marine Science
College of William and Mary
P.O. Box 1346
Cloucester Point, VA 23062
Tel: (804) 684-7105
e-mail: **milliman@vims.edu**

Mark Reagan
Department of Geoscience
University of Iowa
Iowa City, IA 52242
Tel: (319) 335-1802
e-mail: **mark-reagan@uiowa.edu**

Thomas H. Shipley
Institute of Geophysics
University of Texas
4412 Spicewood Springs Road, Building 600
Austin, Texas 78759
Tel: (512) 471-0430
e-mail: **tom@utig.ig.utexas.edu**

Eli A. Silver
University of California, Santa Cruz
Earth Science Department & Institute of Tectonics
1156 High Street
Santa Cruz, California 95064
Tel: (831) 459-2266
e-mail: **esilver@earthsci.ucsc.edu**

Rudy Slingerland
Department of Geosciences
Penn State University
503 Deike Bldg
University Park, PA 16802
Tel: 814-865-6892
e-mail: **sling@geosc.psu.edu**

Joann Stock
Seismological Laboratory 252-21
California Institute of Technology
1200 E. California Blvd.
Pasadena, California 91125
Tel: 626-395-6938
e-mail: **jstock@gps.caltech.edu**

Patricia Wiberg
Department of Environmental Sciences
University of Virginia
P.O. Box 400123
Charlottesville, VA 22904-4123
Tel: (434) 924-7546
e-mail: **pw3c@virginia.edu**

MARGINS Office
Lamont-Doherty Earth Observatory
Oceanography Building, room 305
61 Route 9W, P.O. Box 1000
Palisades, New York 10964-8000
Tel: (845) 365-8665, or 8509
Fax: (845) 365-8156
e-mail: **MARGINS@ldeo.columbia.edu**
or: **olafs@ldeo.columbia.edu**
WWW: http://www.ldeo.columbia.edu/margins

This information is also posted on the MARGINS website, where it is continuously updated.

NSF Program Directors

National Science Foundation, 4201 Wilson Boulevard, Arlington, Virginia 22230

Bilal Haq
Marine Geology and Geophysics Program
Division of Ocean Sciences
Tel: (703) 292-8582
Fax: (703) 292-9085
e-mail: bhaq@nsf.gov

Paul Dauphin
Ocean Drilling Program
Division of Ocean Sciences
Tel: (703) 292-8581
Fax: (703) 292-9085
e-mail: jdauphin@nsf.gov

David Fountain
Tectonics Program
Division of Earth Sciences
Tel: (703) 292-8552
Fax: (703) 292-9025
e-mail: dfountai@nsf.gov

Lamont-Doherty Earth Observatory
of Columbia University

MARGINS

Newsletter No. 9, Fall 2002

Upcoming Meetings:

• AGU Fall Meeting 2002

6-10 December, 2003, San Francisco, Ca

• S2S Town Meeting

7 PM on Saturday, 7 December, 2003, San Francisco, Ca

(see ad on p.38)

• SEIZE Theoretical Institute

16-21 March 2003, Snowbird, Utah

(see ad on p.23)

• Waipaoa Workshop

4-9 May 2003, Gisborne/Wellington, New Zealand

(see ad on p.7)

• AGU~EGS~EUG Joint Assembly

6-11 April 2003, Nice, France

More information about MARGINS-related meetings are posted on the Meetings page at the MARGINS web site:

http://www.ldeo.columbia.edu/margins

MARGINS

NSF

This newsletter is supported by the National Science Foundation under Grant No. OCE 00-79660. Any opinions, findings, and conclusions expressed in it are those of the authors and do not necessarily reflect the views of the National Science Foundation.