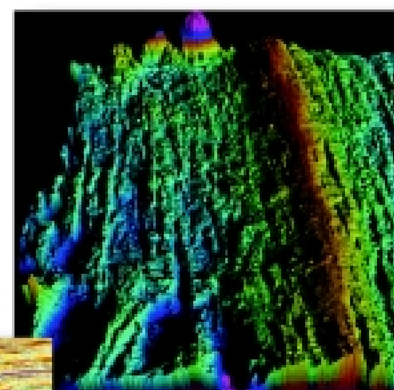
DATA MANAGEMENT FOR MARINE GEOLOGY AND GEOPHYSICS

Tools for Archiving, Analysis, and Visualization



WORKSHOP REPORT
LA JOLLA, CALIFORNIA
MAY 14-16, 2001

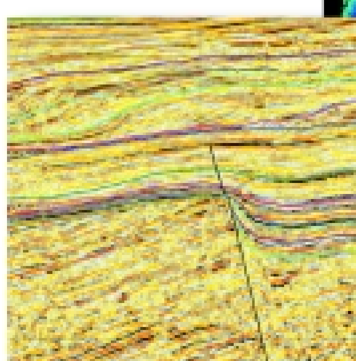


TABLE OF CONTENTS

Support for this workshop was provided by the National Science Foundation and Office of Naval Research.

ORGANIZING COMMITTEE

Steve Cande
SIO/UCSD
9500 Gilman Dr.
La Jolla, CA 92093-0220
cande@gauss.ucsd.edu
Tel: 858-534-1552
Fax: 858-534-0784

Suzanne Carbotte, Co-Chair
LDEO
61 Rte 9W
Palisades, NY 10964
carbotte@ldeo.columbia.edu
Tel: 845-365-8895
Fax: 845-365-8168

Stephen Miller
SIO/UCSD
9500 Gilman Dr.
La Jolla, CA 92093-0220
spmiller@ucsd.edu
Tel: 858-534-1898
Fax: 858-534-0784

William Ryan
LDEO
61 Rte 9W
Palisades, NY 10964
billr@ldeo.columbia.edu
Tel: 845-365-8312
Fax: 845-365-8156

Deborah Smith, Co-Chair
Department of Geology
and Geophysics, MS #22
WHOI
Woods Hole, MA 02543
dsmith@whoi.edu
Tel: 508-289-2472
Fax: 508-457-2187

Dawn Wright
Department of Geosciences
Oregon State University
Corvallis, OR 97331-5506
dawn@dusk.geo.orst.edu
Tel: 541-737-1229
Fax: 541-737-1200

EXECUTIVE SUMMARY

ON MAY 14-16, 2001 the National Science Foundation and the Office of Naval Research sponsored a workshop on Data Management for Marine Geology and Geophysics: Tools for Archiving, Analysis, and Visualization. The workshop was held at the Sea Lodge Hotel in La Jolla, CA. The workshop's objective was to bring together researchers, data collectors, data users, engineers, and computer scientists to assess the state of existing data management efforts in the marine geology and geophysics (MG&G) community, share experiences in developing data management projects, and help determine the direction of future efforts in data management.

The workshop agenda was organized around presentations, plenary discussions, and working group discussions. The presentations provided examples of the needs of data users, the needs of large, multidisciplinary MG&G projects, existing data management projects in the community, tools that have been developed for data access and analysis, examples of organizations with centralized databases, and current topics in information technology.

Working groups addressed questions concerning three different themes: (1) the structure of a data management system, (2) data archiving and access, and (3) data documentation. The working groups were also asked to recommend strategies to permit MG&G data management to move forward in each of these areas.

The Working Groups came up with 20 recommendations:

OVERARCHING RECOMMENDATIONS

1. Create permanent, active archives for all MG&G data.
2. Create a centralized and searchable on-line metadata catalog.

MANAGEMENT SYSTEM

3. Manage data using a distributed system with a central coordination center.
4. Manage different data types with user-defined centers.
5. Support area- or problem-specific databases if scientifically justified, but these databases should link to rather than duplicate data holdings within discipline-specific data centers.
6. Fund core operating costs of the distributed data centers as 3-5 year facility cooperative agreements.
7. Evaluate the data management system using oversight and advisory committees, in-depth peer reviews at renewal intervals, and *ad hoc* panels to assess each data center's contribution to science.

DATA ARCHIVING AND ACCESS

8. Always archive raw data. Archive derived data for high-demand products.
9. Store data in open formats.

10. Develop standardized tools and procedures to ensure quality control at all steps from acquisition through archiving.

11. Improve access to common tools for data analysis and interpretation for the benefit of the community.

12. Build data centers to address the needs of a diverse user community, which will be primarily scientists.

13. Enforce timely data distribution through funding agency actions.

14. Promote interactions among federal agencies and organizations, and international agencies to define data and metadata exchange standards and policies.

DATA DOCUMENTATION

15. Require ship operators and principal investigators (P.I.s) to submit level 1 metadata* and cruise navigation to the centralized metadata catalog at the end of each cruise as part of the cruise-reporting process.

16. Generate a standard digital cruise report form and make it available to all chief scientists for cruise reporting (level 2 metadata*).

17. Require individual P.I.s to complete and submit standard forms for level 1 and 2 metadata* for field programs carried out aboard vessels not in the University-National Laboratory System (UNOLS) fleet (e.g., foreign, commercial, other academic platforms).

18. Generate a standardized suite of level 1 and level 2 metadata* during operation of seafloor observatories and other national facilities (e.g., the Deep Submergence Laboratory, Ocean Bottom Seismograph (OBS) Instrument Pool), and submit to the central metadata catalog.

19. Require level 3 metadata* within each discipline-specific data center. Archiving of publications related to the data should also be included (level 4 metadata*).

20. Follow nationally accepted metadata standards (particularly for levels 1 and 3 metadata*).

A clear top priority of the workshop participants is to immediately define and establish a centralized metadata catalog. The metadata catalog should be broad, containing information on as many data types as possible. It should support geospatial, temporal, keyword, and expert-level searches of each data type. By definition, metadata are information about data that can evolve. The catalog should be a circular system that allows feedback from the user/originator. The metadata catalog should serve as the central link to the distributed network of data centers where the actual data reside.

To move forward, funding agencies must establish a small working group or advisory board to develop the structure and implementation of a metadata catalog. Additional working groups for each of the high-priority, discipline-specific data centers also need to be assembled. It is critical to obtain the active involvement of scientists in all aspects of this process through all operational phases, including data collection, processing, archiving, and distribution.

Section 2 of this report discusses these recommendations further.

*Metadata levels: Level 1. Basic description of the field program including location, program dates, data types, collecting institutions, collecting vessel, and P.I.s. Level 2. A digital cruise report and data inventory. Level 3. Data object and access information including data formats, quality, processing, etc. Level 4. Publications derived from the data.

1. OVERVIEW

1.1 MOTIVATION

MG&G SCIENTIFIC data collections are growing at a rapid rate (Figures 1 and 2). Processed and analyzed data made available to a broad community of scientists, educators, and the general public can be used for discovering and distributing new knowledge. A significant problem is how to provide data users with the means to effectively access these data and the tools to analyze and interpret them.

Advances in data storage technology have eliminated practical constraints on storing large data volumes and have permitted data collection at increasingly finer sample rates. New high-resolution systems provide digital images of the seafloor at sub-meter pixel resolution and generate data at rates on the order of Gigabytes per day (Figures 3 and 4). Seismic acquisition capabilities have greatly expanded with long-term deployment of bottom sensors. High-resolution multichannel seismic reflection (MCS) systems are now available for shallow-water problems, and digital acquisition of 480 channel data is currently routine for deep-ocean work (Figure 5).

With these new technologies it is becoming increasingly difficult for individual investigators to synthesize and organize data sets collected on single cruises let alone manage them in a manner that allows data to be accessed efficiently by a larger user pool. National archiving of some marine geoscience data is carried out. There have been several attempts by individual P.I.s to establish geographic- or data-specific databases. However, access to many data types remains difficult and incomplete (e.g., MCS, multibeam bathymetry, sidescan sonar, camera, and video imagery). Large quantities of data are under-utilized by primary users and gaining access to these data is virtually impossible for secondary users. At the same time, our scientific interests are increasingly interdiscipli-

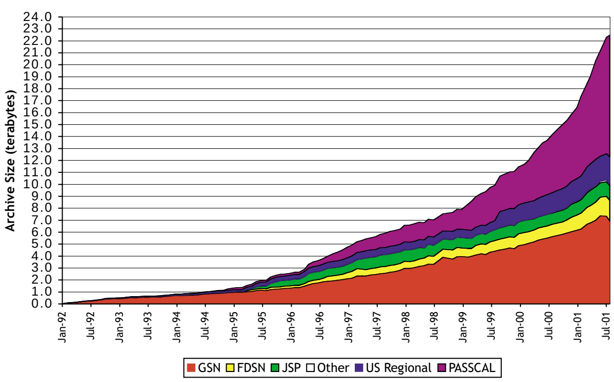


Fig. 1. Growth in the Incorporated Research Institutions for Seismology (IRIS) data archive since 1992. Data holdings from a variety of different seismic networks are shown (GSN – Global Seismographic Network, FDSN – Federation of Digital Broad-Band Seismograph Networks PASSCAL – Program for the Array Seismic Studies of the Continental Lithosphere). Figure provided by Tim Ahern (IRIS).

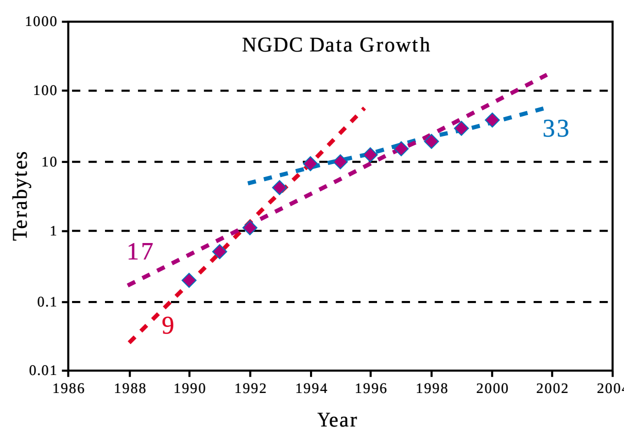


Fig. 2. Growth in the digital data holdings of marine geophysical data at the National Geophysical Data Center (NGDC) since 1990. Dashed lines show data doubling times of 9, 17 and 33 months. Figure provided by George Sharman (NGDC, NOAA).

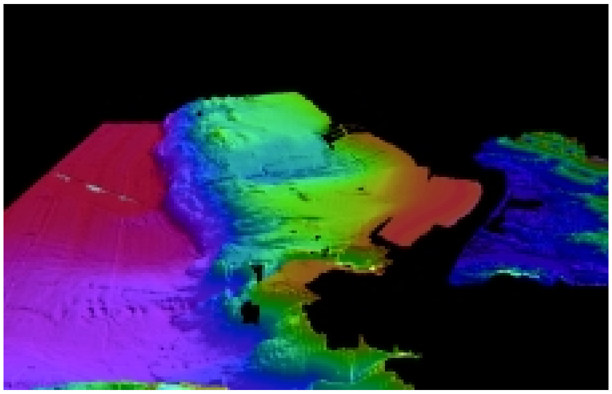


Figure 3. Sun-illuminated perspective view of the Eel River margin of Northern California. Multibeam bathymetry (EM1000, Hydrosweep and Seabeam data) are merged with the USGS 30 m DEM for the adjacent land. Three-dimensional visualization of the merged topography is carried out using Fledermaus from Interactive Visualization Systems and Analysis. (Fonseca, L , Mayer, L. and Paton, M., ArcView Objects in the Fledermaus Interactive 3-D Visualization System: An example from the STRATAFORM GIS, in Wright, D.J. (ed.), Undersea With GIS, Redlands, CA: ESRI Press, in press, 2001).

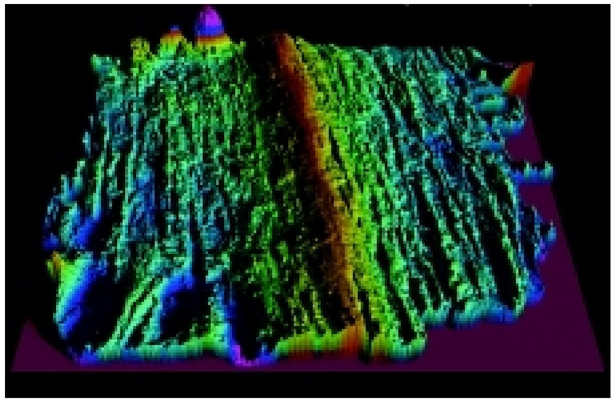


Figure 4. Three-dimensional shaded relief map of the East Pacific Rise near 9º-10ºN. This is currently the best-studied section of fast-spreading mid-ocean ridge. Figure courtesy of Dawn Wright (OSU).

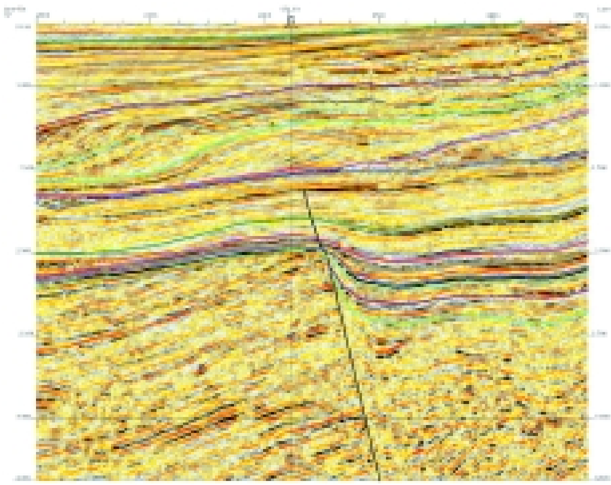


Figure 5. Example of a multichannel seismic reflection record from the northwest shelf of Australia. Seismic interpretation of various reflectors is superimposed with various colors. Data are stored and displayed within the GEOQUEST IESX seismic-well log integrator/ data browser. The power engine of IESX is an Oracle-based database system that organizes seismic, well log and geographical data in a local environment for interpretation. Figure courtesy of Garry Karner (LDEO).

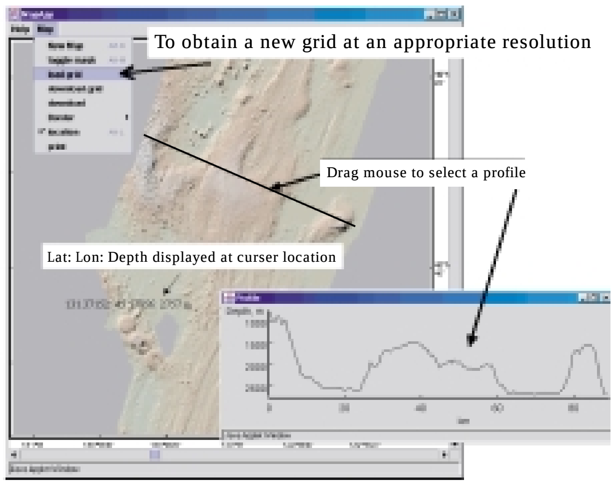


Figure 6 Example of capability provided by MapAp, a web-based map driven, database interface developed for the RIDGE Multibeam Synthesis project (see Appendix 4). Figure shows a multibeam bathymetry map of Axial Seamount, NE Pacific, with a user-defined profile location and corresponding bathymetry profile displayed. Figure provided by Bill Haxby (LDEO).

nary and require easy access to the broad spectrum of data collected. Throughout the marine geoscience community, scientists want access to data, the ability to compare data of different types, and tools to manipulate and interpret these data (Figures 6, 7, 8).

With these concerns in mind, the National Science Foundation (NSF) and Office of Naval Research (ONR) sponsored a workshop on MG&G data management on May 14-16, 2001 in La Jolla, California. The coordinating committee advertised the workshop; participation was open. Approximately 80 representatives from science, engineering, computer science, government, and industry attended (Appendix 1). The workshop provided a forum for a focused interchange of ideas among data users, data providers, and technical experts in information technology.

1.2 OBJECTIVES

The overall goal of the workshop was to develop a strategy for MG&G data management that would meet scientists' needs for improved data access and improved tools for data analysis and interpretation. Accomplishing this goal will lead to greater use of data by the MG&G community and the education and outreach community.

Another workshop objective was to provide NSF and ONR with recommendations on how to implement this data management strategy. The organizing committee thus created three thematic working groups: (1) structure of a data management system, (2) data archiving and access, and (3) data documentation. Each group discussed key problems within their theme and provided a list of recommendations on how to solve them. Critical to implementing these recommendations is active involvement of scientists in the entire process, including collecting, processing, archiving, and distributing data.

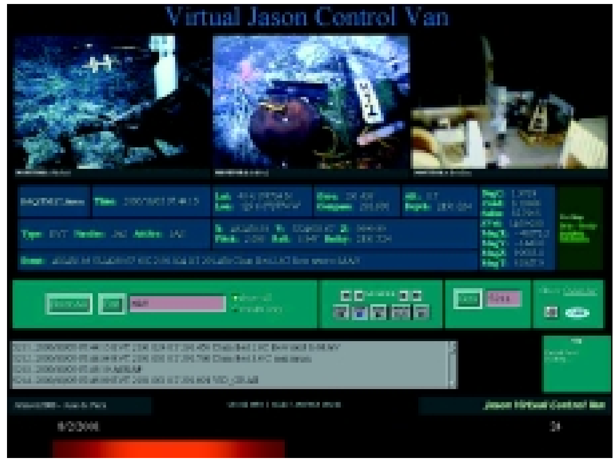


Figure 7. The Virtual Jason Control Van is a web-based application that takes real-time snapshots of information occurring inside the control van during vehicle operations and makes this information immediately available for shipboard scientists and for collaboration and post-cruise analysis on shore. Features include monitoring real-time operations, searching for events, dates, etc. Figure courtesy of Steven Lerner (WHOI).

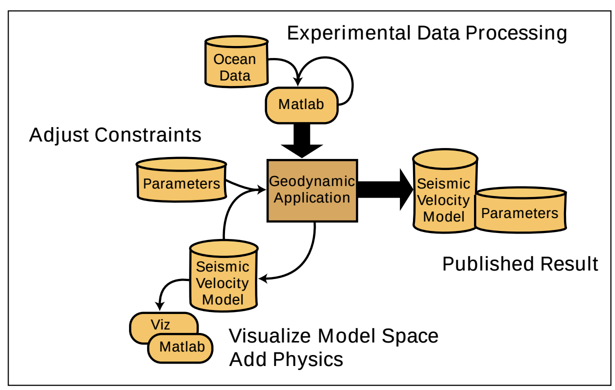


Figure 8. Schematic diagram of scientific workflow. There are three stages that modelers go through when developing a computational result: (1) experimental data processing; (2) parameter adjustment; and (3) publication of the result. It is important that the modeler be able to link tools easily to the output of computational applications (e.g., to visualize the data). Figure courtesy of Dawn Wright (OSU).

1.3 AGENDA

The first day of the workshop was devoted to short talks, each followed by a brief discussion. In addition, a longer discussion followed each group of subject-specific talks. The longer discussion was led by a pre-assigned discussion leader. Our intent was to engage the participants in the meeting right from the start through the discussion.

The workshop began with presentations from data users. The talks focused on problems that P.I.s have had in the past with gaining access to data, and possible solutions to these problems. Representatives from large, multidisciplinary MG&G programs gave overviews on anticipated database needs for new program initiatives. Individual P.I.s made presentations on database projects which they initiated, providing working examples of data access and functionality over a range of disciplines. In the late afternoon of the first day, workshop participants presented models for data access. The format of this session was somewhat different as the talks served as introductions for demonstrations that were part of the evening poster session. In addition to these invited demonstrations, the evening session included posters and demonstrations contributed by workshop participants.

The second day of the workshop began with presentations by representatives of organizations with large central databases. Talks focused on anticipated future directions in data access and database design as well as insights on successes and major obstacles encountered during their efforts to date. The final set of talks focused on current developments in information technology, including data mining issues and designing databases to serve real-time data.

The presentations on days 1 and 2 served as catalysts for discussions that were held within the theme working groups, each of which consisted of an interdisciplinary group of scientists, engineers, and computer scientists. These working groups addressed a number of questions that formed the basis for presentations in the morning of the third day of the workshop.

The full agenda is given in Appendix 2. Presentations and poster abstracts can be obtained through the workshop conveners.

1.4 EVALUATION

An evaluation form was included in the workshop packet that participants received. The forms were collected at the conclusion of the workshop. The responses to the questions have been compiled and are presented in Appendix 3.

1.5 RELEVANT URLS

During the meeting, participants were asked to provide links to web sites that are relevant to MG&G database efforts. This URL list is provided in Appendix 4.

2. WORKING GROUP SUMMARIES

2.1 WORKING GROUP 1

Structure of a Data Management System

Working Group 1 considered how to structure a MG&G data management system. Currently, some data are archived at the National Geophysical Data Center (NGDC), such as the suite of underway geophysical data (navigation, gravity, magnetics, topography) collected on most large UNOLS vessels. However, it is not standard practice to submit to NGDC all data collected by the MG&G community.

Several ship-operating institutions have archived data at some level. However, no standardization across institutions exists and these efforts have been carried out at the discretion of the individual institutions. The need for a sound data management system is recognized, and a few workshops have been held to address this problem for specific data types (e.g., MCS Workshop, La Jolla, CA, 1999). In addition, individual P.I.s have made specific data types available to the broader community (see Appendix 4). It is evident that there is no community-wide strategy in place to solve MG&G data management problems.

QUESTIONS CONSIDERED

Working Group 1 addressed the following questions:

- What model is appropriate for a data management system (e.g., distributed versus centralized)?
- How do we fund the data management system?
- How do we evaluate the system?
- Do we need a permanent archive?

There was clear agreement within the group that the MG&G community needs a distributed data management system with a coordination center to facilitate communication among different data centers. The working group session started with a discussion of metadata, indicating the importance of metadata to attendees and the MG&G community, in general. The consensus of Working Group 1 is that the community must begin taking small, concrete steps towards establishing a metadata catalog. From there the community should move towards a discipline-oriented, distributed data management system that will improve the data use by a broad community. Development of the discipline-oriented data centers should be handled through the normal competitive proposal process. Although participants agreed that significant resources are needed for new database efforts, exact details of the level of government agency funds for the management system were not determined.

RECOMMENDATIONS

WG1_1. Create permanent, active archives for all MG&G data.

It is very important that the funding agencies maintain and strengthen their commitment to long-term data archiving. As noted at the meeting, data collected by the HMS *Challenger* are still being used. Permanent archives for all types of MG&G data must be established. The community must continually add to and update these permanent archives.

WG1_2. Manage data using a distributed system with a central coordinating center.

The management system should operate as close to the data generators as possible. Scientists must be actively involved in data management, placing the responsibility for and authority over the data as close as possible to where the expertise resides. Data quality control should be provided by those generating the data. Mechanisms should be developed to enable users to easily provide feedback on data quality.

A coordination center is necessary to facilitate communications among the distributed data centers, and to ensure that everyone works together. A good example of central coordination is the OBS Instrument Pool. The individual instrument centers provide quality control and write standard format data. The Incorporated Research Institutions for Seismology (IRIS) then archives the data and provides community access to them.

WG1_3. Manage different data types with user-defined centers.

Examples of different data types and their management status are given below. The list is not all inclusive.

1. *Ocean bottom seismograph/hydrophone (OBS/OBH) data.* Quality control is provided by the three OBS instrument centers and archival and community access is provided by IRIS.
2. *Rock petrology/geochemistry data.* A web-served database is being developed to provide metadata and processed results for rock samples. This effort is ready for migration to permanent support.
3. *Core/rock collections.* Sample curation appears to be in good shape. NGDC maintains a central catalog of the existence of physical samples and some sample metadata.
4. *Ocean Drilling Program (ODP) data.* ODP developed the JANUS database based on community recommendations that came out of several workshops. It appears to be in good shape. There are plans in place to transition the database from ODP to IODP in 2003.
5. *Single channel and multichannel seismic data.* A workshop was held in 1999 to determine the needs of the community for database management. Recommendations were made from that workshop. An interested subgroup of the MCS community needs to define the model details and submit a proposal to NSF.
6. *Multibeam sonar data (bathymetry, sidescan, backscatter, LIDAR, etc.).* This community needs a user/generator workshop or working group to define the problems and solutions to their database needs. There is a critical need for one-stop access quality-control and processing centers with tools to generate higher level products. There does not seem to be a quality-control process in place although the MB-System software provides tools for reading a broad suite of multibeam data.
7. *Deep submergence data collected by near-bottom instruments (submarines, remotely operated vehicles (ROVs), autonomous underwater vehicles (AUVs), etc.).* In principle, these data should be managed in the same way that other shipboard data are managed. A data management plan must be defined and overseen, perhaps through the Deep Submergence Science Committee (DESSC), the existing operators and user group.
8. *Gravity/magnetic data.* NGDC maintains archives of these data, but there are major quality-control and user-interface problems. The community concerned about these data needs to be defined. Value-added products (derived products) should be archived and made available to the broad community.
9. *Sedimentology, paleontology data.* Although it was noted that problems exist, there were too few representatives from these communities at the workshop to define the issues and possible solutions. NSF should encourage mini-workshops or working groups for these data.

WG1_4. Support area- or problem-specific databases if scientifically justified, but these databases should link to rather than duplicate data holdings within discipline-specific data centers.

Working Group 1 recognizes that there might be a future need to set up databases for specific oceanic regions or for specific scientific goals. Examples of area-specific databases are those for the 9°N area of the East Pacific Rise and the Juan de Fuca region. Examples of problem-specific databases are those that will develop from the MARGINS and RIDGE programs. These databases should be supported, but they should serve as links to discipline-specific databases and should not duplicate data holdings within these databases.

WG1_5. Evaluate the data management system using oversight and advisory committees, in-depth peer reviews at renewal intervals, and *ad hoc* panels to assess each data center's contribution to science.

The data management system should undergo regularly scheduled peer review. A new set of advisory groups representing the broad spectrum of the MG&G research community should be established. This will ensure that the recommendations regarding data sets and models will be responsive to the community's needs. Selection of data centers should be determined through competition, and a data center should not expect to be funded permanently.

WG1_6. Fund core operating costs of the distributed data centers as 3-5 year facility cooperative agreements.

This is a corollary to recommendation WG1_5 in that funds should cover a finite number of years after which each of the data centers should be evaluated for effectiveness and responsiveness to users' needs.

2.2 WORKING GROUP 2

Data Archiving and Access

Working Group 2 focused primarily on data archiving issues. Problems associated with current MG&G archiving efforts range from complete absence of an archive for many important data types, to lack of quality control and inadequate data delivery to archives. While ship operators deliver underway geophysical data to the NGDC (Figure 9), there are no standards for data quality and it can be difficult to obtain a uniform data set for a region. Multibeam bathymetric systems are currently operated on most deep-ocean vessels within the academic fleet, but standards do not exist for navigation editing, beam editing, or even the nature of the final data product (corrected or uncorrected meters). Some multibeam data are archived with the NGDC, some with the ship operating institutions, and some within problem-specific archives (e.g., the

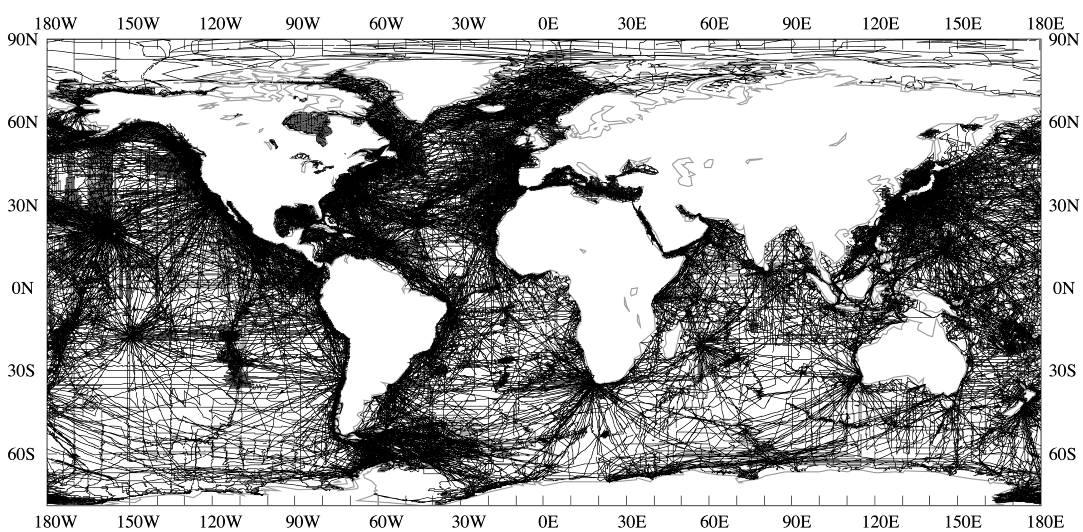


Figure 9. World map showing over 15 million miles of ship tracks with underway geophysical data inventoried within NGDC's Marine Trackline Geophysics Database. Bathymetry, magnetics, gravity, and seismic reflection data along these tracks from 4600 cruises were collected from 1939 to 2000. Figure provided by John Campagnoli, NGDC.

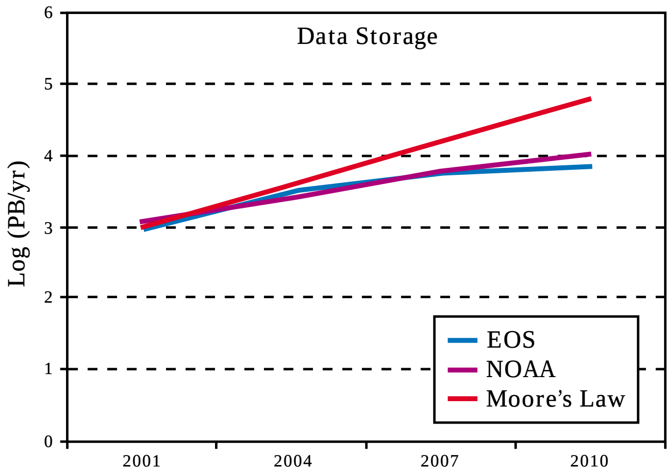


Figure 10. Predicted future growth in data holdings at NGDC compared with data storage capability predicted by Moore's law for the next 10 years. Figure demonstrates that expected data storage capability should be more than adequate to handle expected data volumes. Figure courtesy of Herbert Kroehl, NGDC.

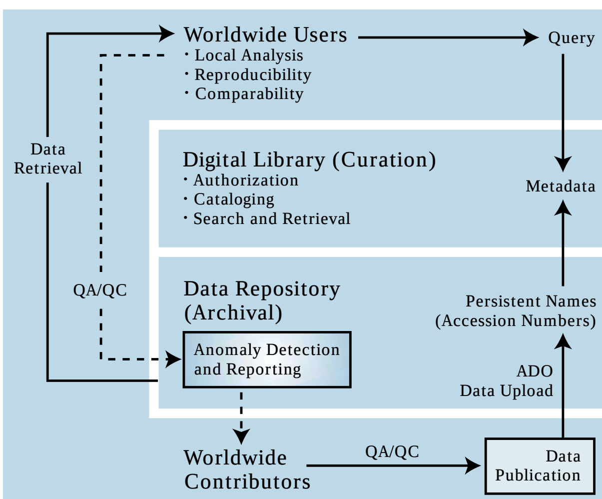


Figure 11. Arbitrary Digital Objects (ADO) are produced when contributed data are uploaded to the data repository. The ADO is assigned a persistent and unique name within the repository. These and other metadata are passed to the digital library function where they can be searched using a catalogue database. Key elements of this process are the continuing involvement of the authors of the data and the maintenance of a dialogue between data users and authors or their successors. Another important consideration is the separation of the metadata catalogue search function from the data repository and delivery function. Both become more portable and reliable when functionally separated. Figure from J. Helly, T. T. Elvins, D. Sutton, D. Martinez, S. Miller, S. Pickett and A. M. Ellison, Controlled Publication of Digital Scientific Data, Communications of the ACM (accepted October 3. 2000).

RIDGE Multibeam Synthesis Web Site). However, delivery to these data archives is largely at the discretion of the P.I., and access to these data and many other data types is often difficult.

Data ownership issues continue to be significant obstacles to archiving efforts. Although NSF policy permits a two-year proprietary hold on data collected by a P.I., data are commonly held well past this time period.

Recognition for contribution to data archives is an important issue that has not been well addressed by any existing archives.

QUESTIONS CONSIDERED

Working Group 2 considered the following questions:

- What data need to be archived?
- Should we archive raw data, processed data and/or interpretations?
- How can we ensure quality control?
- How do we provide broad data access for scientists and the public?
- What tools are needed for data interpretation and analysis?
- How do we enforce timely data distribution?
- How do we reward data contribution to archives?

Overall priorities were defined as: (1) save the data, (2) provide a catalog of all of the data, and (3) provide easy access to the data. With advances in storage technology and web access, the bulk archiving of data is feasible, but our community will be challenged in the areas of quality control and metadata generation (Figures 10 and 11). A centralized metadata catalog of all data-collection activities was viewed by this group as a very high priority. To build this catalog, the working group recommended the development of easy-to-use tools for automatic metadata generation aboard UNOLS vessels. The data types to be archived include, but are not limited to:

- Underway geophysical data, including time, position, magnetics and gravity, multibeam bathymetry, sidescan sonar, single and multichannel seismics.

- Standard supporting data, including sea state, XBT, CTD, sea surface temperature and salinity, and derived sound velocity profiles, as well as calibration data for each sensor.
- Station information for dredging, coring, trawling, and other over-the-side operations, with complete data or metadata, as appropriate.
- Some individual-investigator instrumental data need not be saved, but metadata with time, position, and contact need to be archived. For example, it may not be appropriate to archive test data collected from a prototype sensor, but the existence of these data should be documented and preserved.

RECOMMENDATIONS

WG2_1. Always archive raw data. Archive derived data for high-demand products.

Current data storage capability is adequate for on-the-fly generation of some types of derived data products. However, some types of derived data should be archived for high-demand products such as multibeam bathymetry grids or maps, or when nontrivial processing steps are required (e.g., MCS data).

Easy retrieval of these derived, value-added "products" (e.g., images of reflection profiles, grids, bathymetric maps, graphs) must be developed. Everyone benefits from maximum use of the data including scientists, the government, and the public.

WG2_2. Store data in open formats.

Tested, portable, and noncommercial software for data translation must be freely available for all users.

WG2_3. Develop standardized tools and procedures to ensure quality control at all steps from acquisition through archiving.

Standardized shipboard tools are the first step to ensuring quality control. Easy-to-use, real-time data quality monitoring tools are needed for UNOLS vessels, as well as cost-effective ship-to-shore communication of sufficient compressed data for quality assessment and troubleshooting. Before data are allowed to enter archives, common sanity and geographic-bounds checks need to be applied. Circular archives are needed to permit content to be updated as errors are found by users, with appropriate notations in metadata. The peer-review process in electronic journals can provide broad-based quality assessment, and the publication of data briefs in electronic journals is encouraged.

WG2_4. Improve access to common tools for data analysis and interpretation for the benefit of the community.

A combination of public domain and commercial tools are used widely, including GMT, MB-System, ArcView, Fledermaus, and Matlab. A data center should maintain a list of suitable software for viewing and analyzing each data type, instructions on installation, data exchange and usage for our community, and contacts for further assistance. Custom, open-source development may be needed for special tools and interfaces for community-wide use.

WG2_5. Build data centers with the goal of addressing the needs of a diverse user community, which will be primarily scientists.

Platform-independent, browser-based data extraction tools are needed. The authoritative metadata catalog should support geospatial, temporal, keyword, and expert-level searches of each data type. A federated system of distributed data centers, easily updated and synchronized, will provide reliable and efficient delivery for each data type. Data should be archived in a form easily used by other disciplines. Experience has shown that well-organized, image-rich, searchable databases will serve the needs of both researchers and the public (Figure 12).

WG2_6. Enforce timely data distribution through funding agency actions.

Raw data should be delivered to the designated data center immediately following each cruise. The designated center will restrict data access to the P.I. and identified collaborators for an initial proprietary hold period. This lock is released when the period expires. The standard period is two years, although some circumstances may warrant an extension to be granted by the cognizant funding agency.

Auditing access to data will provide usage statistics and facilitate interdisciplinary collaboration as well as the communication of future updates, within the restrictions of privacy requirements.

The NSF Final Report could include a field to describe how the P.I. complied with NSF data-distribution policies. Noncompliance might have a negative effect on future proposals. Data publication in citable journals and in technical briefs such as USGS open-file reports should be encouraged.

WG2_7. Promote interactions among federal agencies and organizations, and international agencies to define data and metadata exchange standards and policies.

The community would benefit from the standardization of forms, such as an end-of-cruise digital data form, as well as from metadata content and exchange standards. We encourage collaboration among the federally mandated agencies (NSF, ONR, USGS, NOAA, NAVO, etc.) to review marine database standards.

International discussions should be encouraged to define exchange standards and policies. At a minimum, exchange of cruise tracks and sample locations would be a major benefit for cruise planning.

2.3 WORKING GROUP 3
Data Documentation

Working Group 3 focused on metadata issues. The development of appropriate metadata and metadata standards for ocean floor and other types of oceanographic data are an extremely important issue. The growth in information technology has led to an explosion in the amount of information that is available to researchers in many fields. This is the case in the marine environment where a state-of-the-art “visual presence” (e.g., through long-term monitoring by cameras and other instruments) may result in the acquisition of data that quickly overtakes the speed at which the data can be interpreted. The paradox is that as the amount of potentially useful and important data grows, it becomes increasingly difficult to know what data exist, the exact location where the data were collected (particularly when navigating at sea with no “landmarks”), and how the data can be accessed. In striving to manage this ever-increasing amount of data, and to facilitate their effective and efficient use, compiling metadata becomes an urgent issue.

Although metadata are contained within some of the digital data file formats commonly used to store MG&G data (e.g,. MGD77 and SEGY formats), no uniform metadata are collected during federally funded MG&G field programs. Basic information regarding cruise location, date, project P.I.s, and data types collected can be difficult to obtain, and no central and comprehensive catalog is available. Cruise reports often contain detailed information regarding general experiment configuration, data calibration, and data quality, all of which are of great importance for subsequent data analysis. In many instances, the cruise report may be the only record of this information, but no easily accessible digital archive of these reports exists.

QUESTIONS CONSIDERED

Working Group 3 addressed the following questions:

- How do we move toward metadata standards?
- How do we standardize data collection procedures?
- What is the role of the ship operating institutions in the archiving of data and generation of metadata?
- What existing software and structures should we take advantage of?

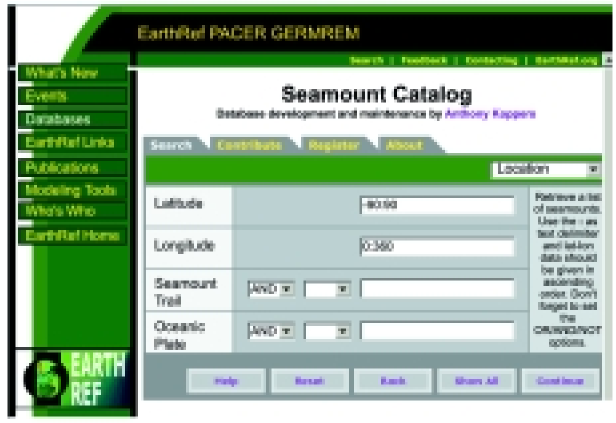
- How should we deal with real-time data acquisition?

To aid the development of a standardized procedure for generating metadata during MG&G studies, four levels of metadata were defined, each defining a particular stage of the data-acquisition to publication process:

- Level 1. Basic description of the field program including location, program dates, data types, collecting institutions, collecting vessel, and P.I.s.
- Level 2. A digital cruise report and data inventory.
- Level 3. Data object and access information including data formats, quality, processing, etc.
- Level 4. Publications derived from the data.

Responsibility for each metadata level could reside with different groups (e.g., ship-operating institution or P.I.) and some metadata generation could be automated and standardized across UNOLS vessels. The construction of a central metadata catalog for levels 1 and 2 metadata was viewed as the highest priority. The group consensus is that level 1 metadata should be generated during data acquisition and should be submitted to the central metadata archive immediately following a field program. Level 2 metadata should also be archived within the central metadata catalog, whereas level 3 metadata would reside with the actual data themselves. The appropriate archive for level 4 metadata may be both the central metadata catalog and the individual data centers. The requirements for levels 1 and 2 metadata should be standardized whereas level 3 requirements will vary by data type.

The group's consensus is that a first step toward a central metadata catalog is to develop and implement procedures for metadata collection for all future MG&G data-acquisition efforts. Archiving and rescue efforts for legacy data should be handled as a parallel but secondary priority and should begin with cataloging existing data.



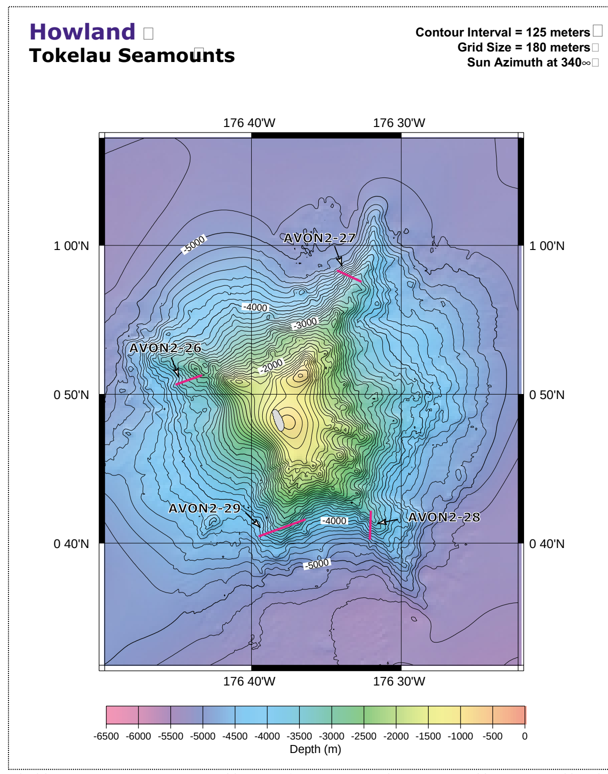


Figure 12. Example of an on-line database where the interface permits users to search for data available from seamounts. Bottom figure shows a contoured bathymetry map for Howland seamount. Figure courtesy of Anthony Koopers and Stephen Miller (SIO).

RECOMMENDATIONS

WG3_1. Create a centralized and searchable on-line metadata catalog.

The metadata catalog should be broad, containing information on as many data types as possible. It should support geospatial, temporal, keyword, and expert-level searches of each data type. By definition, metadata are information about data that can evolve. The catalog should be a circular system that allows feedback from the user/originator. The metadata catalog should serve as the central link to the distributed network of data centers where the actual data reside.

Selection of an organization to develop and maintain this metadata catalog should be through a competitive process. The organization will oversee the development of metadata entry tools for easy entry into the metadata catalog. A high performance storage system to archive and serve the catalog to the community should also be implemented.

WG3_2. Require ship operators and P.I.s to submit level 1 metadata and cruise navigation to the centralized metadata catalog at the end of each cruise as part of the cruise reporting process.

This function should be provided by the technical support staff aboard UNOLS vessels, although the ultimate responsibility for generating and delivering these data should lie with the project P.I. Tools need to be developed to facilitate this task, simplifying the process with a smart web form. Standard forms should be used on all UNOLS vessels and for all kinds of data-collection activities (chemical, physical, biological, and geological studies). Level 1 metadata along with cruise navigation should be submitted.

UNOLS may be an appropriate organization to manage the metadata submission process (and possibly the catalog), perhaps through modification of the UNOLS electronic ship request form. Metadata need to be defined, but should include items such as the chief scientist(s), project P.I.(s), institution(s), data types collected, dates of field program, geographic coordinates of the field area, ship name, and cruise leg ID (if appropriate). Metadata standardization is very important. Metadata and data need to be handled separately for maximum efficiency.

WG3_3. Generate a standard digital cruise report form and make it available to all chief scientists for cruise reporting (level 2 metadata).

These digital forms should be uniform across all federal agencies for all future cruises and should be submitted to the centralized metadata catalog.

Old cruise reports should be digitized, perhaps from the NOAA National Oceanographic Data Center (NODC) archive, as a parallel effort. Standard reporting should include essential fields described above as well as specific details for each data type (e.g., data ranges for each data type, acquisition quality control records, number and location of sample stations). The responsible individual and physical location where each data type will reside following a cruise should be identified.

WG3_4. Require individual P.I.s to complete and submit standard forms for level 1 and 2 metadata for field programs carried out aboard non-UNOLS vessels (e.g., foreign, commercial, other academic platforms).

Not all field programs carried out by MG&G researchers involve UNOLS vessels, and procedures need to be developed that permit the cataloging of data collected during these programs as well.

WG3_5. Generate a standardized suite of level 1 and 2 metadata during operation of seafloor observatories as well as other national facilities (e.g., the Deep Submergence Laboratory, OBS Instrument Pool) and submit to the central metadata catalog.

The metadata required should parallel that acquired from UNOLS operations with additional fields as relevant. Navigation from submersibles, ROVs, and AUVs needs to be captured and archived along with support-ship navigation.

WG3_6. Require level 3 metadata within each discipline-specific data center.

Required metadata for a specific data type will likely vary and will be decided through development of individual data centers. These metadata include, for example, descriptions of data formats, retrieval information, data quality, and processing procedures. Archiving of publications related to the data should also be included (level 4 metadata).

WG3_7. Follow nationally accepted metadata standards (particularly for metadata levels 1 and 3).

A national content standard for metadata has already been established by the Federal Geographic Data Committee (FGDC). The standard is being migrated to match international ISO metadata standards, and fully outlines as much vital information as possible pertaining to a data set's source, content, format, accuracy, and lineage (i.e., what processing changes the data set has gone through over time). The content standard was developed by the FGDC primarily for GIS and satellite remote-sensing data as one way of implementing the National Spatial Data Infrastructure (NSDI).

We recommend taking advantage of these efforts. The FGDC standard is extremely complex, but small portions of it will be very useful in the creation of workable metadata standards for the various subdisciplines of MG&G.

APPENDIX 1: LIST OF ATTENDEES

May 14-16, 2001

NAME	AFFILIATION	E-MAIL
Ahern, Tim -SPEAKER	IRIS	tim@iris.washington.edu
Arko, Robert	LDEO	arko@ldeo.columbia.edu
Barclay, Andrew	Univ. of Washington	abarclay@whoi.edu
Bartling, William	SciFrame, Inc.	wbartling@sciframe.com
Baru, Chaitan	UCSD	baru@sdsc.edu
Batiza, Rodey	NSF	rbatiza.nsf.gov
Bemis, Karen	Rutgers Univ.	bemis@rci.rutgers.edu
Benson, Rick	IRIS	rick@iris.washington.edu
Bobbitt, Andra -SPEAKER	OSU	bobbitt@pmel.noaa.gov
Brovey, Robert	Exxon/Mobil	rlbrove@upstream.xomcorp.com
Cande, Steve -CONVENER	SIO/UCSD	scande@ucsd.edu
Carbotte, Suzanne -CONVENER	LDEO	carbotte@ldeo.columbia.edu
Caress, David	MBARI	caress@mbari.org
Carron, Michael	Naval Oceanogr. Office	carronm@navo.navy.mil
Case, James D.	SAIC	jcase@mtg.saic.com
Chayes, Dale	LDEO	dale@ldeo.columbia.edu
Childs, Jonathan	USGS	jchilds@usgs.gov
Christie, David - SPEAKER	OSU	dchristie@oce.orst.edu
Cochran, James	LDEO	jrc@ldeo.columbia.edu
Cushing, Judy -SPEAKER	Evergreen State	Judith.Cushing@inria.fr
Domenico, Ben -SPEAKER	Unidata/UCAR	ben@unidata.ucar.edu
Eakins, Barry	SIO/UCSD	beakins@ucsd.edu
Epp, David	NSF	depp@nsf.gov
Fornari, Daniel -DISCUSSION LEADER	WHOI	dfornari@whoi.edu
Gahagan, Lisa	UTIG	lisa@utig.ig.utexas.edu
Gaudet, Severin –SPEAKER	Canadian Astr. Data Center	Severin.Gaudet@nrc.ca
Gourley, Mike -SPEAKER	CARIS	Gourley@caris.com
Griffin, Nicole	Univ. of New Orleans	ngriffin@uno.edu
Habermann, Ted -SPEAKER	NOAA/NGDC	Ted.Habermann@noaa.gov
Heath, G. Ross	Univ. of Washington	rheath@u.washington.edu
Helly, John -SPEAKER	UCSD	hellyj@ucsd.edu
Henkart, Paul	SIO/IGPP	phenkart@ucsd.edu
Hey, Richard	HIGP/SOEST	hey@soest.hawaii.edu
Howland, Jonathan	WHOI	jhowland@whoi.edu
Jenkins, Chris	Univ. of Sydney	cjenkins@es.usyd.edu.au
Karner, Garry -SPEAKER	LDEO	garry@ldeo.columbia.edu
Kent, Graham	SIO/IGPP	gkent@ucsd.edu
Knoop, Peter	Univ. of Michigan	knoop@umich.edu
Koppers, Anthony	SIO/IGPP	akoppers@ucsd.edu
Kroehl, Herb -SPEAKER	NGDC	hkroehl@ngdc.noaa.gov

NAME	AFFILIATION	E-MAIL
Kurras, Greg	Univ. of Hawaii/ HMRG	gkurras@soest.hawaii.edu
Langmuir, Charlie -SPEAKER	LDEO	langmuir@ldeo.columbia.edu
Lawrence, Richard -SPEAKER	ESRI	rlawrence@esri.com
Le Bas, Tim	SOC	tlb@soc.soton.ac.uk
Lehnert, Kerstin	LDEO	lehnert@ldeo.columbia.edu
Lenhardt, W. Christopher	CIESIN	clenhardt@ciesin.columbia.edu
Lerner, Steve -SPEAKER	WHOI	slerner@whoi.edu
Lonsdale, Peter -SPEAKER	SIO/UCSD	plonsdale@ucsd.edu
Mayer, Larry -SPEAKER	Univ. of New Hampshire	lmayer@unh.edu
McCann, Mike	MBARI	mccann@mbari.org
McDuff, Russell	Univ. of Washington	mcduff@ocean.washington.edu
Miller, Stephen -CONVENER	SIO/UCSD	spmiller@ucsd.edu
Moore, Reagan	UCSD	moore@sdsc.edu
Morris, Peter	British Antarctic Survey	pmor@bas.ac.uk
Mountain, Gregory	LDEO	mountain@ldeo.columbia.edu
Naar, David	Univ. of South Florida	naar@usf.edu
Orcutt, John -DISCUSSION LEADER	SIO/IGPP	jorcutt@ucsd.edu
Rack, Frank -SPEAKER	JOI-ODP	frack@brook.edu
Reagan, Mary	LDEO	mreagan@ldeo.columbia. edu
Reid, Jane A.	USGS	jareid@usgs.gov
Ryan, William -CONVENER	LDEO	billr@ldeo.columbia.edu
Sarbas, Barbel	Max-Planck Inst.	sarbas@mpch-mainz.mpg.de
Scheirer, Daniel	Brown Univ.	scheirer@emma.geo.brown.edu
Sharman, George -SPEAKER	NGDC	gsharman@ngdc.noaa.gov
Shaw, Kevin B.	NRL	shaw@nrlssc.navy.mil
Shipley, Tom -SPEAKER	UTIG	tom@ig.utexas.edu
Small, Chris -SPEAKER	LDEO	small@ldeo.columbia.edu
Smith, Deborah -CONVENER	WHOI	dsmith@whoi.edu
Smith, Stu	SIO/UCSD	ssmith@ucsd.edu
Staudigel, Hubert	SIO/IGPP	hstaudigel@ucsd.edu
Su, Young	LDEO	ysu@ldeo.columbia.edu
Syvitski, James	Univ. of Colorado	james.syvitski@colorado.edu
Tivey, Maurice	WHOI	mtivey@whoi.edu
Tolstoy, Maya	LDEO	tolstoy@ldeo.columbia.edu
Toomey, Douglas	Univ. of Oregon	drt@newberry.uoregon.edu
Vaquero, Maria	Spanish Oceanogr. Inst.	juan.acosta@md.ieo.es
Wessel, Paul	Univ. of Hawaii	pwessel@hawaii.edu
Wilcock, William	Univ. of Washington	wilcock@ocean.washington.edu
Wright, Dawn -CONVENER	OSU	dawn@dusk.geo.orst.edu
Zaslavsky, Ilya	UCSD	zaslavsk@sdsc.edu

APPENDIX 2: FINAL AGENDA

DAY 1: MONDAY MAY 14

7:00 am – Continental Breakfast

8:00 am – 8:15 am – Plenary Session - Introduction to the Workshop

8:15 am – 9:50 am - Data user needs (15-minute talks, 10 minutes for questions)

Presentations from individual data users.

Tom Shipley (UTIG)

Peter Lonsdale (SIO)

Chris Small (LDEO)

Discussion – 20 minutes, DL: Dan Fornari (WHOI)

9:50 am – 10:20 am Break

10:20 am – 11:55 am - Large programs (15-minute talks, 10 minutes for questions)

Presentations from representatives of large programs.

Dave Christie (OSU) - RIDGE

Gary Karner (LDEO) - MARGINS

Severin Gaudet (Canadian Astronomy Data Center) - NEPTUNE

Discussion – 20 minutes DL: John Orcutt (SIO/IGPP)

11:55 am – 12:20 pm –Existing projects, P.I. driven (15-minute talks, 10 minutes for questions)

Presentations on individual P.I.-initiated data management projects.

Suzanne Carbotte/Bill Ryan (LDEO)

12:20 pm - 1:30 pm Lunch

1:30 pm – 3:30 pm – Existing Projects P.I. driven, continued (15-minute talks, 10 minutes for questions)

Charlie Langmuir (LDEO)

Larry Mayer (UNH)

Andra Bobbitt (NOAA/PMEL)

Dawn Wright (OSU)/Judy Cushing (Evergreen State College)

Discussion – 20 minutes - DL: Dawn Wright (OSU)

3:30 pm – 4:00 pm Break

4:00 pm – 5:00 pm – Tools for data access and analysis (10-minutes talks, 5 minutes for questions). Presentations on models for data access.

Ted Habermann (NGDC)

Richard Lawrence (ESRI)

Steve Lerner (WHOI)

Mike Gourley (CARIS)

5:00 pm – 5:15 pm - Summary of Day 1

6:00 pm: Reception/Poster Session at SIO/IGPP Munk Conference Room.

DAY 2: TUESDAY, MAY 15

7:00 am – Continental Breakfast

8:00 am – 8:15 am – Plenary Session - Introduction to Day 2

8:15 am – 10:15 am - Organizations with centralized databases (15-minute talks, 10 minutes for questions) Presentations on large database efforts.

Frank Rack (ODP)

Tim Ahern (IRIS)

George Sharman (NGDC)

Stephen Miller (SIO)

Discussion – 20 minutes; DL: Steve Cande (SIO)

10:15 am – 10:45 am Break

10:45 am – 12:20 pm – Database components (15-minute talks, 10 minutes for questions)

Presentations on information technology issues.

Herb Kroehl (NOAA)

John Helly (UCSD)

Ben Domenico (UCAR)

Discussion – 20 minutes, DL: Severin Gaudet (Canadian Astronomy Data Center)

12:20 pm – 1:30 pm Lunch

1:30 pm – 1:45 pm – Plenary session – Define goals of the Working Groups

1:45 pm - 3:15 pm - Working Groups

Break into multidisciplinary groups to address questions

3:15 pm – 3:45 pm Break

3:45 pm – 5:15 pm - Working Groups

5:15 pm – 5:30 pm – Summary of Day 2

Evening: Dinner on your own. Tour of San Diego Supercomputer Center

DAY 3: WEDNESDAY, MAY 16

7:00 am – Continental Breakfast

8:15 am – 10:00 am - Plenary Session

Working Group summaries – break into Working Groups if needed.

10:00 am – 10:30 am Break

10:30 pm – 12:00 pm - Plenary Session

Where do we go from here? List recommendations.

12:00 pm

End of meeting

APPENDIX 3: WORKSHOP EVALUATION

The following workshop evaluation consists of answers to two questions and a list of additional comments made by the participants. The evaluation was collected from the participants at the end of the workshop. Following the comments by the participants, pie diagrams of the session evaluation data are presented. There were 35 forms submitted (~ 45% of the participants). Not all participants answered each question.

QUESTION 1: Was there adequate time for each activity?

Yes - 28 No - 5

If you were of the opinion there was inadequate time, please explain.

- The working groups required much more time to discuss their issues. Also, the size (too large) prohibited focused discussions.
- More time for working groups.
- Too many issues that are unfamiliar to the majority of participants were brought up and a final recommendation is premature. More meetings with focus groups seem required.
- The size of the meeting was too large – too many people. The time required is proportional to the square of the number of attendees (2x people need 4x time).
- The time for "tools for data access and analysis" was a little bit limited. It is understandable that the time allotted for commercial presentations was less than the other general presentations but the time allotted did not allow for much interaction with the audience.

QUESTION 2: What single suggestion would you make to improve this workshop?

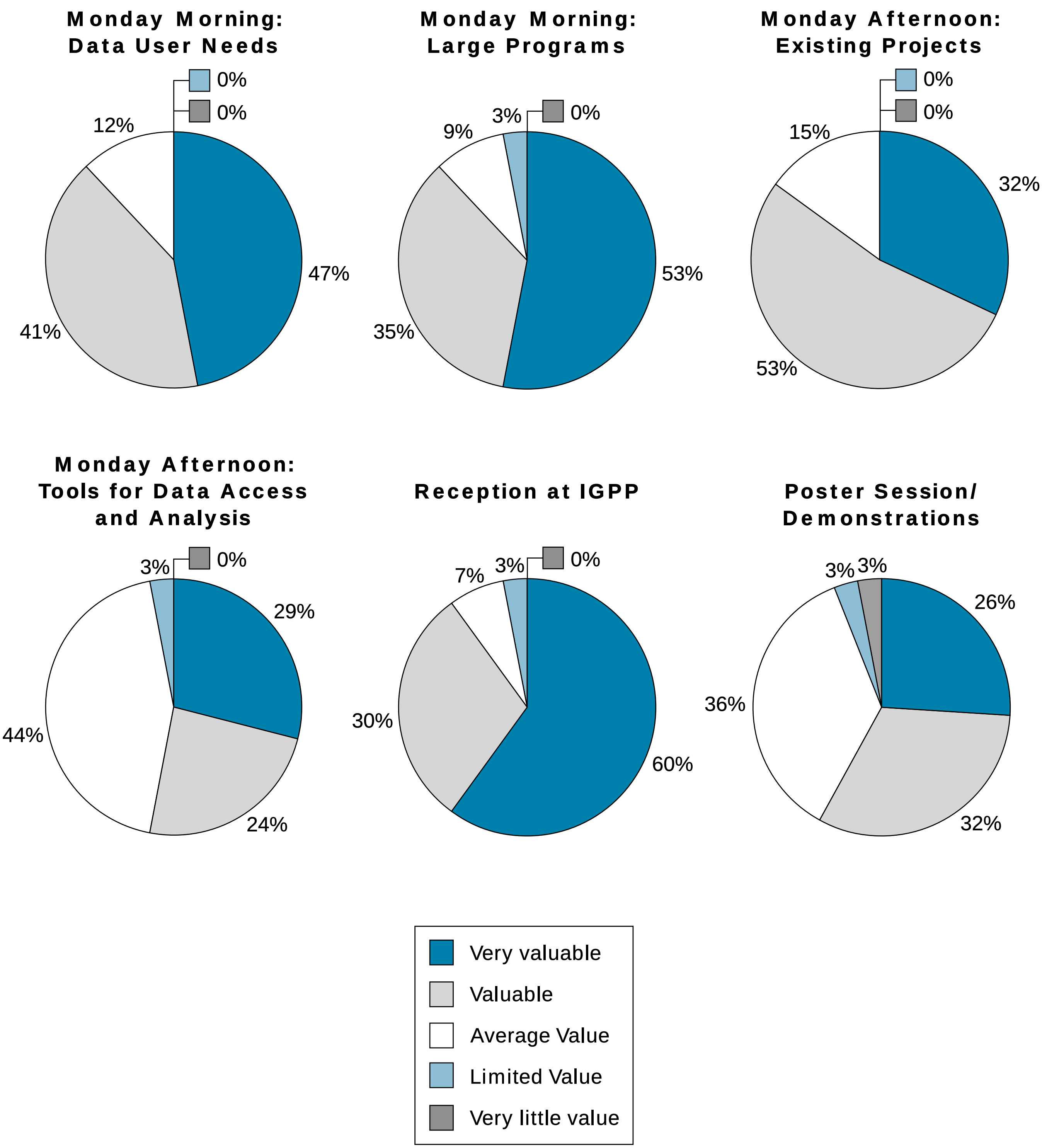
- Needed an example of a working data information system on the WWW (such as land use system).
- Follow up with another one in a year or so.
- Would have been nice to have more input from funding agencies.
- Discussion was dominated by data providers. Clear visions of what the long term goals "should" be often got lost. Long term goals should have more user input, including general nonscientific community.
- Mandates to working groups were somewhat vague and overlapping. Need to be more focused and carefully thought out.
- Present proposals prior to workshop—maybe developed by very small groups.
- Reconvene at least once within 12-18 mo. after the proceedings and recommendations have been disseminated and reviewed by the NSF management and community.
- Ask NSF PMs to talk about NSF commitment to workshop objectives at the end of the workshop.
- Better fit of room to audience size.
- None.
- None.
- A room that would make it easier to see the presentations. However, the surroundings were pleasant and the location at the hotel was convenient.

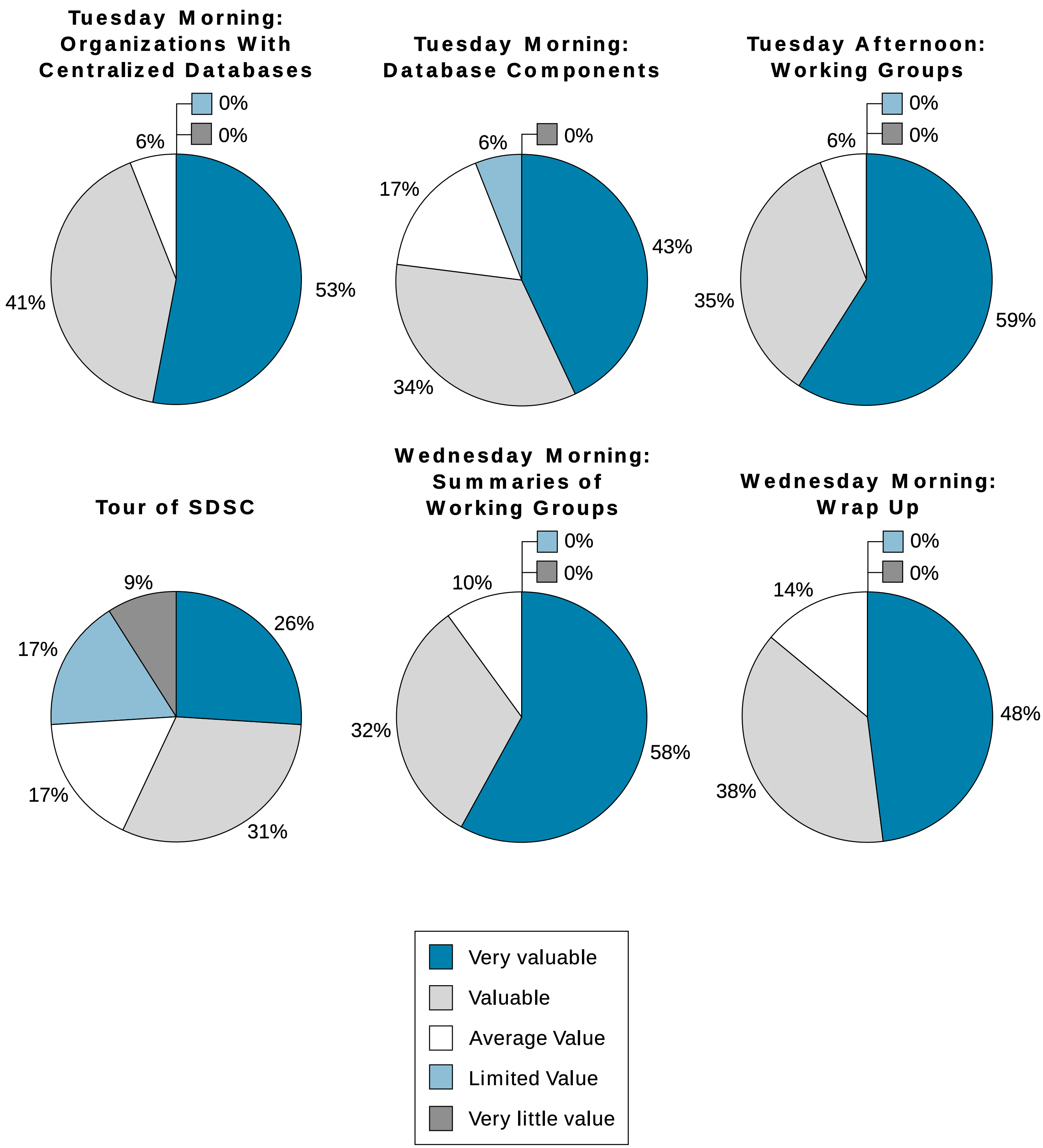
- Provide a summary of existing workshop recommendations on database management in other fields.
- More UNOLS participation (especially since we generated "unfunded mandates").
- Invite international attendees like the French.
- Organize it so that it focuses on more specific recommendations, less general (?).
- A more focused group of experts from both the scientific and computer science communities should be gathered to improve progress, where domain experts in data base management and science plan a detailed proposal to NSF.
- Small item but short description of agenda items would be useful.
- Handout of overheads/presentation slides.
- None.
- The sessions probably should have had a mandate to develop some themes or recommendations and the session leaders could have been given the mandate to develop some consensus or themes as part of the session. These "results" could have been fed into the working groups to make them more productive.
- None.
- Abstracts and titles of talks available before meeting.
- There should have been "read-ahead" material to inform participants about other database discussions and workshops that have already taken place under NSF sponsorship.
- I thought having the people from "outside" the MG&G community (esp., Cushing, Gaudet, Brovey) was a good idea. Perhaps a bit more input from the oil industry would have been good —they collect very similar data and face similar problems: serving up data, what media to store data on.
- More IT (Information Technology) specifics.
- It was a good balance of researchers and workers associated with DB systems. Job well done.
- Fewer, more select audience/participants at the risk of compromised broadness to achieve a higher degree of focus.
- More pre-meeting planning and distribution of material.
- Internet access at the meeting! Posters at the meeting!
- More info provided prior to the meeting.
- More focus on who and how this is going to make this happen.

ADDITIONAL COMMENTS

- Room should be laid out broad and shallow instead of long and deep.
- The follow-up workshop should emphasize more focused groups of users and providers by discipline on data type (e.g., MB, MCS, UW Video, etc.)
- Very well run workshop.
- My background is in C.S. and I enjoyed this conference very much!!
- Very informative – learned a lot of what is done and available.
- I hope that CARIS would be invited back.
- Thanks for the big effort to organize it.
- The meeting was exceptionally well choreographed and no problems with the time lines.
- This was a useful fact-finding workshop, but the details on how the future will be mapped is not clear at the end of this session.
- Very useful workshop. Good to see consensus building throughout. More productive than many workshops.
- I think the workshop was very successful in gathering the experience and articulating the needs and concerns of the MG&G community. The key will be to craft recommendations that will lead to coherent actions.
- I thought Gaudet's talk on the data he worked with (the amount and flow) was good as it put the amount of data our group is discussing into perspective. It gives me a sense that we should be able to organize the data that we have.
- An important and refreshing opportunity to rethink and reconsider NGDC/MGGs role and responsibilities to the community.
- Data catalog vs. database distinction is important. I would like to have seen more examples of working solutions such as the one that Peter Knoop presented.

SUMMATIVE EVALUATION DATA




Tuesday Morning: Organizations With Centralized Databases
0%
0%
6%
41%
53%
Tuesday Morning: Database Components
6%
0%
17%
43%
34%
Tuesday Afternoon: Working Groups
0%
0%
6%
35%
59%
Tour of SDSC
9%
26%
17%
17%
31%
Wednesday Morning: Summaries of Working Groups
0%
0%
10%
32%
58%
Wednesday Morning: Wrap Up
0%
0%
14%
48%
38%
Very valuable
Valuable
Average Value
Limited Value
Very little value

APPENDIX 4: RELEVANT URLS

PARTICIPANT-PROVIDED WEB SITES

Institution	Site	Description
CARIS	http://www.caris.com	Marine software solutions, MB/SB/SSS processing, Hydrographic database
CARIS	http://www.spatialcomponents.com	Spatial Fusion web mapping
CIESIN	http://sedac.ciesin.columbia.edu/gateway	Distributed metadata catalog search tool
CIESIN	http:// sedac.ciesin.columbia.edu/plue/gpw	Gridded population of the world
ESRI	http://www.esri.com	GIS, ArcInfo, ArcView, ArcIMS
IRIS	http://iris.washington.edu	Iris data management center
LDEO	http://www.ldeo.columbia.edu/adgrav	Antarctic digital gravity synthesis
LDEO	http://www.ldeo.columbia.edu/cgif	Coastal geophysics imaging facility
LDEO	http://www.ldeo.columbia.edu/ ~ dale/dataflow	Discussion of bits to data
LDEO	http://coast.ldeo.columbia.edu	Ridge multibeam synthesis
LDEO	http://petdb.ldeo.columbia.edu	Ridge petrological database
LDEO	http://www.ldeo.columbia.edu/SCICEX	SCICEX, SCAMP Arctic mapping with US Navy submarines
Max-Planck Institut fur Chemie in Mainz	http://georoc.mpch-mainz.gwdg.de	Geochemical database for oceanic island, island arcs, LIPs
MBARI	http://www.mbari.org/data/mapping	MBARI multibeam and other data
NOAA/NGDC	http://www.ngdc.noaa.gov/mgg/mggd.html	National archives of underway, multibeam, and seafloor sediment/rock data
NOAA	http://www.pmel.noaa.gov/vents/data	NOAA Vents program data gateway
NOAA	http://www.pmel.noaa.gov/vents/acoustics.html	Underwater Acoustic Monitoring
NOAA	http://newport.pmel.noaa.gov/nemo/realtime/	NeMO Net Real-Time Monitoring & Data
ODP/LDEO	http://www.ldeo.columbia.edu/BRG/ODP/DATABASE	ODP downhole logging database
ODP	http://www-odp.tamu.edu/database	Janus database
Oregon St./ U of Oregon	http://dusk.geo.orst.edu/vrv http://www.cs.uregon.edu/research/vrv-et	Virtual Research Vessel, MOR data access and online computational environment
Oregon St	http://dusk.geo.orst.edu/djl	Davey Jones' Locker seafloor mapping and marine GIS
Oregon St	http://buccaneer.geo.orst.edu/dawn/tonga	Boomerang 8 cruise database (Tonga trench and forearc)
SAIC	http://www.oe.saic.com	Ocean Explorers – Public domain dataset rapid visualization tool
SIO	http://gdcmp1.ucsd.edu	Geological data center cruise archives
SIO	http://www.earthref.org	Geochemical Earth Reference Model (GERM), seamount catalogue
SIO	http://sioseis.ucsd.edu/reflection_archive	Seismic reflection data
SIO	http://topex.ucsd.edu/marine_topo/mar_topo.html	Global Topography
SOEST	http://www.soest.hawaii.edu/STAG/data.html	Marine data archives
SOEST	http://www.soest.hawaii.edu/HMRG	HMRG data archives
SOEST	http://ahanemo2.whoi.edu/	AHA-NEMO 2 cruise database
UCAR	http://www.unidata.ucar.edu	Tools for accessing and visualizing real time data
U Michigan	http://www.si.umich.edu/SPARC	Space Physics and Aeronomy Research Collaboratory – web-based access to distributed databases
U Sidney	http://www.es.usyd.edu.au/geology/centres/osi/ auseabed/au7_web.html	Australian seabed database

USGS	http://walrus.wr.usgs.gov/infobank	Coastal and marine metadatabank
USGS	http://edc.cr.usgs.gov	Land surface data, DEM's, satellite imagery
UTIG	http://www.ig.utexas.edu/srws	UTIG Seismic reflection data holdings
UW	http://bromide.ocean.washington.edu/gis/	Endeavour Segment GIS
WHOI	http://4dgeo.whoi.edu/virtualvan	Jason Virtual Control Van
WHOI	http://science.whoi.edu/kn16213	Recent cruise database for ROV Jason cruise to the Indian Ocean
WHOI	http://drifor.whoi.edu/LuckyStrike96/	MAR Lucky Strike database
WHOI	http://www.divediscover.whoi.edu	Dive and Discover public cruise outreach
WHOI	http://mbdata.whoi.edu/mbdata.html	Multibeam data archives

RELATED SITES – ORGANIZATIONS AND DATABASES

Provided by Steve Miller (SIO) and Dawn Wright (OSU)

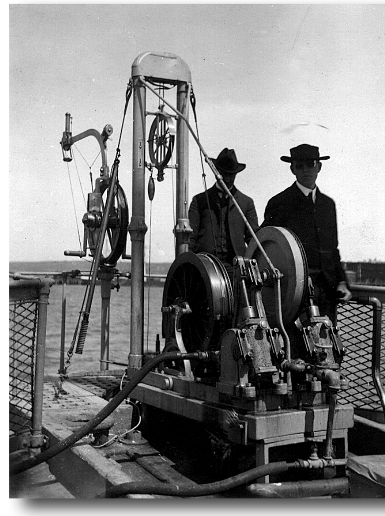
Acronym	Institution	Site	Description
	AGSO	http://www.agso.gov.au/databases http://www.agso.gov.au/marine/marine.html	Multibeam surveys of Australian Territorial waters: Marine and Coastal Data Directory: GEOMET metadatabase
ADEPT	UCSB	http://www.alexandria.ucsb.edu/	Alexandria Digital Earth Prototype (ADEPT)
DLESE	DLESE	http://www.dlese.org http://www.dlese.org/Metadata	Digital Library for Earth System Education
	Geological Survey of Ireland	http://www.gsiseabed.ie	Database and chart management for surveys of entire coastal zone
	GEOMAR	http://www.geomar.de/projekte/alle/expedition.html	Cruise imagery and publications
GOMaP	NRL/UW	http://www.neptune.washington.edu/pub/documents/gomap_pilot.html	Global Ocean Mapping Program
	IFREMER	http://www.ifremer.fr/sismer/sismer/serveura.htm	SISMER Oceanographic data Center
NBII	NSF	http://www.nbii.gov/	National Biological Information Infrastructure
NEEDS		http://www.needs.org/ http://www.synthesis.org/	National Engineering Education Delivery System
	NIWA/LINZ	http://www.niwa.cri.nz/NIWA_research/coastal.html	New Zealand territorial waters surveys and databases
	NOAA	http://www.csc.noaa.gov/opis	NOAA Ocean GIS (Southeast U.S.)
NPACI	NPACI	http://www.npaci.edu/ http://www.npaci.edu/About_NPACI/index.html	National Partnership for Advanced Computing Infrastructure
NSDL	NSF	http://www.ehr.nsf.gov/due/programs/nsdl/ http://www.smete.org/nsdl/	National SMETE (Science, Mathematics, Engineering and Technology Education) Digital Library
	OSU	http://buccaneer.geo.orst.edu	Oregon Coast Geospatial Clearinghouse
SOPAC	SIO	http://sopac.ucsd.edu	SOPAC geodetic archives and GIS
THREDDS	UCAR	http://www.unidata.ucar.edu/projects/THREDDS/Overview/Home.htm	Thematic Realtime Earth Data Distributed Servers
UCGIS	UCGIS	http://www.ucgis.org	University Consortium for Geographic Information Science
	USGS/Microsoft	http://www-nmd.usgs.gov/esic/esic.html http://microsoft.terraserver.com	"Digital Backyard" - USGS & Microsoft TerraServer
WOCE		http://www-ocean.tamu.edu/WOCE/uswoce.html http://whpo.ucsd.edu	World Ocean Circulation Experiment

TOOLS FOR DATA ACCESS AND ANALYSIS

Provided by Stephen Miller (SIO)

Tool	Organization	Site	Description
ArcGMT	Oregon St	http://dusk.geo.orst.edu/arcgmt	GMT < -- > GIS converter
ArcInfo ArcIMS ArcView	ESRI	http://www.esri.com	GIS, Web GIS
CARIS	CARIS	http://www.caris.com	Marine software solutions, MB/SB/SSS processing, Hydrographic database
DICE	SDSC	www.npaci.edu/DICE/	Data Intensive computing, technology development and software toolkits
ERMapper	Earth Resource Mapping	www.ermapper.com	Image mapping and manipulation
FGDC	FGDC	http://www.fgdc.gov	Federal Geographic Data Committee (Metadata, Clearinghouses)
Fledermaus	IVS	http://www.ivs.unb.ca	3D visualization, QA
Geomedia	Intergraph	http://www.intergraph.com	GIS, format translator, web server
Geoshare	Geoshare	http://www.geoshare.com	Nonprofit corporation for managing exchange of petroleum related data and software
GMT	SOEST	http://www.soest.hawaii.edu/gmt/	Generic Mapping Toolkit
Macromedia Dreamweaver	Macromedia	http://www.macromedia.com	Web content development
MATLAB	Matlab	http://www.mathworks.com	Modeling, display, analysis
MBSYSTEM	MBARI	http://www.mbari.org/~caress/MB-System_intro.html	Multibeam System Seafloor mapping toolkit –public domain
MrSID	LizardTech	www.lizardtech.com	Multi-resolution Seamless Image Database
Open GIS	Open GIS Consortium	http://opengis.opengis.org/wmt/	GIS standards and techniques, Web Mapping Testbed
Oracle	Oracle	http://www.oracle.com http://oai,oracle,com/pls/oai_site/oai_site.home	Database management, web serving
SRB	SDSC	http://www.npaci.edu/online/v5.4srb118.html http://www.npaci.edu/DICE/SRB/	Storage Resource Broker
	UNB/Ocean Mapping Group	http://www.omg.unb.ca/omg/	Software tools for swath bathymetry and sidescan sonar
	UNH/Center for Coastal and Ocean Mapping	http://www.ccom-jhc.unh.edu/	Multibeam processing, statistical beampoint editing, GIS operability

Photos on the back cover are courtesy of the SIO Archives (top right) and Emerson Hiller Personal Collection, Woods Hole Oceanographic Institution, Data Library and Archives (left middle, bottom center).



September 2001

Editing and design by
Geosciences Professional Services, Inc.

Support for this workshop was provided
by the National Science Foundation
and Office of Naval Research.