Developing a Community Approach to

Data Integration, Publication and Archival (DIPA)

Fall AGU Meeting 2002

Program for Saturday 7 December 2002

❖ Data Integration, Publication, and Archival (DIPA) I

OS61C : Saturday Morning 08:30 in MCC 274

❖ Data Integration, Publication, and Archival (DIPA) II: Posters

OS62B : Saturday Afternoon 13:30 in MCC Hall D

❖ Magnetic Database Developments: Public Forum

GP62B : Saturday Afternoon 16:15 in MCC 121

Table of Contents

Session Summary

Developing a Community Approach to Data Integration, Publication and Archival (DIPA)

Ocean sciences are experiencing an exponential growth of data from a wide range of measurement platforms including from satellites, buoys, ships or submersibles. Data integration, publication and archival (DIPA) of these data is one of the major challenges in ocean sciences today.

The purpose of the session will be establish a scientific vision for data integration, publication and archival, and to enumerate and describe existing efforts at organizing data collections and review protocols for geoscience data interchange and continue the discussion among members with respect to growing a federation of data collections in the geosciences. From this basis we will identify opportunities for standardization via draft protocols and standards that will be published in appropriate journals for community comment, review and adoption. Emphasis will be placed on commonalities across data collections as well as addressing specific unique requirements of individual data collections.

This session will help define a vision for DIPA, establish an inventory for which groups contribute to these problems in which ways, and define where the major new developments need to happen.

Conveners:

John Helly, Tom Jordan, Dale Chayes, Chaitan Baru, and Hubert Staudigel.

About EarthRef.org

EarthRef.org is an outgrowth of the (geophysical) Reference Earth Model (REM) and the Geochemical Earth Reference Model (GERM) Initiatives. One of the key mandates of EarthRef.org is to strengthen the relationship between Earth sciences and Information Technology, through dialogues at meetings like the DIPA session at AGU, and through a website that supports the linkage between Earth sciences and IT.

Posting of this volume and additional DIPA materials

EarthRef.org has summarized the abstracts of the DIPA session in this volume, that is available as a downloadable PDF at EarthRef.org at **http://earthref.org/cgi-bin/erda.cgi?n=136**. This volume will be archived by EarthRef.org to provide an easily accessible record of this meeting, as it has done for previous EarthRef.org activities, in particular, for the GERM initiative. In addition, EarthRef.org will give authors the opportunity to post additional materials that relate to their abstracts. Such materials may include additional text, illustrations, powerpoint presentations, or entire posters. Please send these materials as e-mail attachments to **pkeizer@ucsd.edu** using PDF, PowerPoint or Adobe Illustrator.

Free GERM Beer at the VGP reception following the DIPA session

Saturday, 7 December from 5:45 - 7:00 P.M. in Moscone Center, Room 132.

The *next* GERM Workshop
will be hosted by F. Albarede and J. Blichert-Toft in Lyon, May 20-23, 2003.

Sign up for the mailing list at **http://earthref.org/events/GERM/2003/**
Field trip reservation at **http://www.bocuse.com/services/rese_aub.html**

Data Base Initiatives of MARGINS

See also the data base special volume of the MARGINS Initiative with a variety of articles on data base efforts related to the MARGINS Initiative, and the philosophy of data base development.
http://www.ldeo.columbia.edu/margins/MARGINS_Newsletter9.pdf

Program

❖ Data Integration, Publication, and Archival (DIPA) I

Saturday Morning 08:30 in MCC 274

Time	Session	Title
0830	**OS61C**	**Presiding: Helly, J; Chayes, D**
0835	OS61C-01 INVITED	Data Integration Across the Geoscience Disciplines: Challenges and Opportunities **Snyder, W S**
0850	OS61C-02 INVITED	Data Collection and Distribution within the IRIS Data Management System: Embracing New Technologies **Ahern, T K**; Benson, R B
0905	OS61C-03	User-Friendly Data Servers for Climate Studies at the Asia-Pacific Data-Research Center (APDRC) **Yuan, G**; Shen, Y; Zhang, Y; Merrill, R; Waseda, T; Mitsudera, H; Hacker, P
0920	OS61C-04	GERM in EarthRef.org: A reference model approach to data bases **Staudigel, H**; Koppers, A A P; Constable, C; Helly, J
0935	OS61C-05	New Magnetic Database Initiatives: Exploitation of and Integration with other Developments **Constable, C**; Staudigel, H; Tauxe, L; Koppers, A A P; Johnson, C; Solheid, P; Jackson, M; Banerjee, S; Pisarevsky, S
0950	OS61C-06	SIOExplorer: Overview, Initial Results and Next Steps **Miller, S P**; Helly, J; Africa, M; Peckman, U; Day, D; Clark, D
1005	OS61C-07 INVITED	Processing, Archiving, and Disseminating Large Swath Mapping Datasets Using MB-System **Caress, D W**; Chayes, D N
1040	OS61C-08	The Evolution of Global Oceanic Crust From Jurassic to Present Day: A Global Data Integration **Mueller, R D**; Gaina, C; Roest, W R; Clark, S; Sdrolias, M
1055	OS61C-09 INVITED	Interoperability Among Spatial Data Resources Along a Continuum: "Data-to-Data", "Data Models" and "Data-to-Interpretation" **Wright, D J**
1110	OS61C-10	A design for the geoinformatics system **Allison, M L**
1125	OS61C-11	Designing Extensible Data Management for Ocean Observatories, Platforms, and Devices **Graybeal, J**; Gomes, K; McCann, M; Schlining, B; Schramm, R; Wilkin, D
1140	OS61C-12	Viewing and Editing Earth Science Metadata MOBE: Metadata Object Browser and Editor in Java **Chase, A**; Helly, J
1155	OS61C-13	The CompreHensive collaborativE Framework (CHEF) **Knoop, P A**; Hardin, J; Killeen, T; Middleton, D

❖ Data Integration, Publication, and Archival (DIPA) II: Posters

Saturday Afternoon 13:30 in MCC Hall D

Time	Session	Title
1330	**OS62B**	**Presiding: Constable, C; Helly, J**
1330	OS62B-0244	Virtual Oregon: A Proof-of-Concept for Seamless Access to Distributed Environmental Information **Keon, D**; Pancake, C; Wright, D J; Walsh, K
1330	OS62B-0245	New Challenges in Sample-Based Data Implementation **Lehnert, K A**; Lenhardt, C; Langmuir, C H
1330	OS62B-0246	The Integration and Enhancement of Seafloor and Land Topography With Satellite Data **Brovey, B**
1330	OS62B-0247	Implementation and compatibility of a North American Volcanic and Plutonic rock database (NAVDAT) **Walker, J D**; Glazner, A F; Bowers, T D; Straus, A K; Farmer, G L; Carlson, R W; Black, R A; Grossman, J N
1330	OS62B-0248	Earth Science Markup Language: An Overview **Ramachandran, R**; Conover, H; Graves, S
1330	OS62B-0249	Using GeoVRML for Visual Dissemination of Oceanographic Data **McCann, M P**
1330	OS62B-0250	Video data annotation, archiving, and access **Wilkin, D**; Connor, J; Stout, N J; Walz, K; Schlining, K; Graybeal, J
1330	OS62B-0251 INVITED	Real-time Metadata Capture Implementations **Chayes, D N**; Arko, R A
1330	OS62B-0252	A Web-Based Geospatial Metadata Browser **Arko, R A**; Chayes, D N
1330	OS62B-0253	Scalable Models of Data Sharing in the Earth Sciences **Helly, J J**; Staudigel, H; Koppers, A A P
1330	OS62B-0254	CruiseViewer: SIOExplorer Graphical Interface to Metadata and Archives **Sutton, D W**; Helly, J J; Miller, S P; Chase, A; Clark, D
1330	OS62B-0255	SIOExplorer: Managing Data Flow into a Digital Library **Clark, D**; Miller, S P; Peckman, U; Chase, A; Helly, J
1330	OS62B-0256	NASA's Eos ClearingHOuse: Integrating Access to Data Services **Burnett, M T**; Pfister, R; Wichman, K
1330	OS62B-0257	Potential of Scalable Vector Graphics (SVG) for Ocean Science Research **Sears, J R**
1330	OS62B-0258 INVITED	A metadata scheme for a rock-magnetic data base - considerations and applications **Dekkers, M J**; Peters, C; Jackson, M J; Solheid, P
1330	OS62B-0259 INVITED	Visual Palaeomagnetic Database **Pisarevsky, S**

1330	OS62B-0260	PMAG: Database Development Under the EarthRef.org Umbrella Website **Koppers, A A P**; Constable, C; Tauxe, L;, Staudigel, H; Helly, J
1330	OS62B-0261	PMAG: Relational Database Definition **Keizer, P**; Koppers, A A P; Tauxe, L; Constable, C; Genevey, A; Staudigel, H; Helly, J
1330	OS62B-0262	Paleointensity Database : Current State and Future **Perrin, M**
1330	OS62B-0263	The Global Paleomagnetic Database Should Include Specimen-Level Demagnetization Data **Kirschvink, J L**; Kazansky, A
1330	OS62B-0264	Archival and Retrieval of Multi-Dimensional Rock Magnetic Data: A Job for BLOB's **Solheid, P A**; Kamp, W P; Jackson, M J; Marvin, J A; Banerjee, S K
1330	OS62B-0265	PMAG: Database Examples for Paleomagnetic and Archeomagnetic Studies **Genevey, A**; Koppers, A A P; Tauxe, L; Constable, C
1330	OS62B-0266 INVITED	Integral Interpretation of Rock Magnetic, Sedimentological and Geochemical Data in Marine Sediment Studies **von Dobeneck, T**; Funk, J
1330	OS62B-0267	Performance Evaluation of INMARSAT Fleet 77 Services Aboard the R/V Ewing **Schmidt, V E**; Chayes, D N; Gold, E

❖ Magnetic Database Developments: Public Forum

Saterday Afternoon 16:15 in MCC 121

Time	Session	Title
1615	**GP62B**	**Presiding: Banerjee, S**
1615	OS62B	Panelists: Van der Voo, R: University of Michigan; Helly, J J: San Diego Supercomputer Center; Bloxham, J: Harvard University; Solheid, P: University of Minnesota; Tauxe, L: Scripps Institution of Oceanography; Fabian, K: University of Bremen; Constable, C: Scripps Institution of Oceanography; Laj, C: CNRS

Abstracts

listing in alphabetical order

OS61C-02; INVITED; EOS, Trans. AGU, 83 (47), F698

Data Collection and Distribution within the IRIS Data Management System: Embracing New Technologies

Ahern, T K; Benson, R B
tim@iris.washington.edu
IRIS DMC, 1408 NE 45th Street Suite 201, Seattle, WA 98105 United States

The IRIS Data Management Center in Seattle has been instrumental in addressing the needs of a global community as it relates to the collection and distribution of seismological data. We will focus on approaches that support distributed models of data management that leverage local expertise and resources. We will offer examples of how effective archiving can be accomplished when standards are first adopted and accepted by a large international scientific community, and what the time scale was to accomplish this. We will offer suggestions on topics such as how to manage real time data streams, managing restricted data, how to insure proper credit is given to data providers and data versioning. We will discuss our current Information Technology (IT) initiatives that directly relate to distributed data access, including observational time series and metadata. We will summarize the concept of Networked Data Centers (NetDC), which connects globally distributed data centers with common interface utilities and eliminates the need to know where data are archived. We will also highlight the FISSURES initiative that includes the Data Handling Interface (DHI), a comprehensive effort to leverage industry-standard CORBA technology to standardize the interfaces to information in distributed data centers. http://www.iris.washington.edu

[back to program schedule]

OS61C-10; EOS, Trans. AGU, 83 (47), F700

A design for the geoinformatics system

Allison, M L
lallison@kgs.ukans.edu
Kansas Geological Survey, 1930 Constant Avenue, Lawrence, KS 66047-3726 United States

Informatics integrates and applies information technologies with scientific and technical disciplines. A geoinformatics system targets the spatially based sciences. The system is not a master database, but will collect pertinent information from disparate databases distributed around the world. Seamless interoperability of databases promises quantum leaps in productivity not only for scientific researchers but also for many areas of society including business and government. The system will incorporate: acquisition of analog and digital legacy data; efficient information and data retrieval mechanisms (via data mining and web services); accessibility to and application of visualization, analysis, and modeling capabilities; online workspace, software, and tutorials; GIS; integration with online scientific journal aggregates and digital libraries; access to real time data collection and dissemination; user-defined automatic notification and quality control filtering for selection of new resources; and application to field techniques such as mapping. In practical terms, such a system will provide the ability to gather data over the Web from a variety of distributed sources, regardless of computer operating systems, database formats, and servers. Search engines will gather data about any geographic location, above, on, or below ground, covering any geologic time, and at any scale or detail. A distributed network of digital geolibraries can archive permanent copies of databases at risk of being discontinued and those that continue to be maintained by the data authors. The geoinformatics system will generate results from widely distributed sources to function as a dynamic data network. Instead of posting a variety of pre-made tables, charts, or maps based on static databases, the interactive dynamic system creates these products on the fly, each time an inquiry is made, using the latest information in the appropriate databases. Thus, in the dynamic system, a map generated today may differ from one created yesterday and one to be created tomorrow, because the databases used to make it are constantly (and sometimes automatically) being updated.

[back to program schedule]

OS62B-0252; EOS, Trans. AGU, 83 (47), F707 (POSTER)

A Web-Based Geospatial Metadata Browser

Arko, R A; Chayes, D N
arko@ldeo.columbia.edu
Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, NY 10964 United States

We are developing a simple Web-based browser for the search and display of earth science metadata. Our design goals are: 1. to permit both map-based (geographical) and forms-based (textual) searching; 2. to integrate a wide variety of data types in a hierarchical fashion; 3. to conform to the FGDC metadata standard; 4. to take advantage of existing open source software wherever possible; 5. to be platform-independent, browser-independent, and "robust" (i.e. avoid application layers which are resource-intensive or behave unpredictably, such as Java applets); and 6. to present metadata in a dynamic fashion via live database connections. Our implementation is based on the MapServer GIS platform (developed at the University of Minnesota with NSF and NASA funding), PostgreSQL relational database management system, and PostGIS geographic database extensions (developed by Refractions Research Inc and available under GNU Public License). All of these packages are well-documented open source software and have been proven in commercial-grade applications. We combine geographical searching (click-and-drag on maps, in both global and polar projections) and textual searching (drop-down menus organized by FGDC category) for a range of geophysical, chemical, and biological data types. A corresponding framework for collecting and ingesting earth science metadata is reported elsewhere at this meeting (Chayes \& Arko, "Real-time Metadata Capture Implementations"). http://data.ldeo.columbia.edu/

[back to program schedule]

OS62B-0246; EOS, Trans. AGU, 83 (47), F706 (POSTER)

The Integration and Enhancement of Seafloor and Land Topography With Satellite Data

Brovey, B
rlbrove@upstream.xomcorp.com
ExxonMobil Upstream Research Co., P.0. Box 2189, Houston, TX 77252-2189 United States

This presentation will illustrate several useful techniques to integrate disparate marine and land data. Six data sets were combined, so that a variety of geophysical information is represented while preserving spatial detail through the entire elevation range. There is significant utility in bringing these data together in a single map/image. The color component of the image depicts bathymetry in the marine setting and vegetation distribution on the land areas. The colors of the marine portion of the image were derived from Smith and Sandwell's bathymetric solution below 72 degrees of latitude and ETOPO5 data above 72 degrees of latitude. The color is scaled to accentuate bathymetric features, such as the continental shelf and mid-ocean ridges. The color component of the land was enhanced from an AVHRR false color infrared dataset that was acquired from WorldSat International. These colors on the marine and land data sets were draped over the textural component of the image. A key objective of this map was to preserve as much of the tectonic detail as possible for the complete bathymetric and topographic range. The marine textural information was derived from Sandwell and Smith's bathymetric and gravity solutions. The land texture came from GTOPO30 data. The texture was differentially enhanced to accentuate areas of low relief. The image is artificially illuminated from the north. For nearly 20 years ExxonMobil has had the opportunity to integrate large grids of marine data received from many sources. This is a recent image that follows our earlier efforts using data acquired from Sandwell, Smith, Rapp, Haxby, The USGS, The World Data Center, and others. Significant improvements in the quality and utility of marine data have been very achieved in part to the efforts and cooperation of many academic and government organizations.

[back to program schedule]

OS62B-0256; EOS, Trans. AGU, 83 (47), F708 (POSTER)

NASA's Eos ClearingHOuse: Integrating Access to Data Services

Burnett, M T
mburnett@blueprinttech.com
Blueprint Technologies, 8618 Westwood Center Suite 310, Vienna, VA 22182 United States
Pfister, R
NASA-ESDIS, Bldg 32, Rm S224C Mail Code 423 Code 586 NASA Goddard Space Flight Center , Greenbelt, MD 20771 United States
Wichman, K
GST, 7855 Walker Drive, Greenbelt, MD 20771 United States

ECHO (The Earth Observing System (EOS) ClearingHOuse) is being developed to provide flexibility to NASA's EOS to better meet the needs of the science community. ECHO is a clearinghouse of metadata, representing the data offerings of participating data providers. ECHO is being built with the goal of being an enabling system: Enabling a variety of Data Providers to participate. Enabling access to an ever-changing variety of Earth Science Data. Enabling access to an ever-growing suite of services, provided by the Science Community, which improves the usefulness of this data, including the binding of those services to the data represented in the clearinghouse. The purpose of this enabling philosophy is to support current Science efforts, but also to give the opportunity for creative organizations and individuals to break the traditional paradigm for discovering and leveraging Earth Science Data and Services in completely new ways. This presentation will focus on ECHO's approach to integrating Data Services from varied Service Providers, and facilitating access to those services by the user community. ECHO can be viewed as a typical Service oriented architecture. The fundamental interactions that it supports are (abstractly) Publish, Find and Bind. ECHO provides interfaces and mechanisms that allow organizations to publish their services. Using these interfaces, Service Providers can effectively "plug-in" their capabilities. There are mechanisms that allow the correlation of their service to the data types in the clearinghouse. ECHO's user community can find, or discover, services through a separate set of interfaces. Bindings are the mechanisms that support the invocation of services by ECHO's user community. ECHO supports binding either directly between the user and the service provider, or indirectly by using ECHO as a Service Broker. ECHO is supporting all of these Service capabilities by leveraging the contemporary (and evolving) "standards" of Web Services. Web Services are implemented using a suite of technologies, all based on eXtensible Markup Language (XML). Services present their interfaces (API's) by using Simple Object Access Protocol (SOAP). The Services are described in Web Services Description Language (WSDL). Finally, ECHO will manage a registry of these Services by leveraging Universal Description, Discovery and Integration (UDDI). Using these technologies offers us tremendous potential, and solutions to many of the challenges associated with a distributed service architecture. However, there are several issues related to the use of this approach in general, and each of the technologies specifically. It is recognized that while providing an API-based system is the only way to retain the principle of being an enabling system, people need User Interfaces with which to interact. ECHO provides an approach to allow organizations to offer User Interfaces that correlate to registered services. ECHO is currently working with a number of potential Service Providers to exercise this approach to enabling access to distributed Data Services. This interaction allows us to improve the mechanisms while beginning to offer Services to the Science community in the short term.

[back to program schedule]

OS61C-07; INVITED; EOS, Trans. AGU, 83 (47), F699

Processing, Archiving, and Disseminating Large Swath Mapping Datasets Using MB-System

Caress, D W
caress@mbari.org
Monterey Bay Aquarium Research Institute, 7700 Sandholdt Rd., Moss Landing, CA 95039 United States
Chayes, D N
Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, NY 10964 United States

MB-System is an NSF and MBARI-funded open source software package for the processing and display of swath mapping sonar data. Version 5.0 of MB-System is structured to enable the management of large datasets. The new software integrates editing and analysis tools with a single program, mbprocess, which outputs processed data files. This parallel approach now allows processing to proceed in a more flexible, efficient fashion. An additional important benefit is that the data management structure allows the active processing environment to be embedded within the overall data archive. Within this structure, bathymetry grids, sidescan mosaics, maps, images, GIS layers, and other data products are generated using the most up-to-date processed data. If one layers data dissemination tools and environments on top

of the data archive, the served data products will also automatically reflect the latest processing efforts. MBARI, L-DEO, and NOAA-NOS are currently developing web-served swath data archives that use MB-System for processing, data product generation, and low-level data management. The aims of these efforts vary, and are reflected in differing high-level data server architectures. We will present and discuss the current state of these swath data archives. http://www.ldeo.columbia.edu/MB-System

[back to program schedule]

OS61C-12; EOS, Trans. AGU, 83 (47), F700

Viewing and Editing Earth Science Metadata MOBE: Metadata Object Browser and Editor in Java

Chase, A; Helly, J
achase@sdsc.edu
San Diego Supercomputer Center, University of California, San Diego 9500 Gilman Drive , La Jolla, CA 92024 United States

Metadata is an important, yet often neglected aspect of successful archival efforts. However, to generate robust, useful metadata is often a time consuming and tedious task. We have been approaching this problem from two directions: first by automating metadata creation, pulling from known sources of data, and in addition, what this (paper/poster?) details, developing friendly software for human interaction with the metadata. MOBE and COBE(Metadata Object Browser and Editor, and Canonical Object Browser and Editor respectively), are Java applications for editing and viewing metadata and digital objects. MOBE has already been designed and deployed, currently being integrated into other areas of the SIOExplorer project. COBE is in the design and development stage, being created with the same considerations in mind as those for MOBE. Metadata creation, viewing, data object creation, and data object viewing, when taken on a small scale are all relatively simple tasks. Computer science however, has an infamous reputation for transforming the simple into complex. As a system scales upwards to become more robust, new features arise and additional functionality is added to the software being written to manage the system. The software that emerges from such an evolution, though powerful, is often complex and difficult to use. With MOBE the focus is on a tool that does a small number of tasks very well. The result has been an application that enables users to manipulate metadata in an intuitive and effective way. This allows for a tool that serves its purpose without introducing additional cognitive load onto the user, an end goal we continue to pursue. http://sioexplorer.ucsd.edu

[back to program schedule]

OS62B-0251; INVITED; EOS, Trans. AGU, 83 (47), F707 (POSTER)

Real-time Metadata Capture Implementations

Chayes, D N; Arko, R A
dale@ldeo.columbia.edu
Lamont-Doherty Earth Observatory of Columbia University, Instrument Lab 61 Route 9W, Paliades, NY 10964 United States

The current rate of data acquisition in the ocean sciences precludes the manual generation of appropriate metadata after the fact. Recognizing this fact, we have begun to implement methods for creating metadata and inserting them into relational databases in real-time. We have also created web-based tools for watchstanders and maintenance personnel to enter logbook data in real-time. Several examples will be addressed in this poster. Enhancements to the Hudson Interactive River Observatory (HIRO) real-time data logging system have been made that create metadata records and insert them (as SQL transactions over a secure wireless TCP/IP connection) into a relational database in real-time. These records document the start and stop time of individual data files, of sensor-specific data streams and of the logging system as a whole. An interactive watchstanders logbook has been developed and used on the R/V Maurice Ewing to create and log metadata records associated with upgrades to the Hydrosweep DS2 multibeam system. A similar version of this tool is being used to capture the maintenance and update records associated with the HRIO system. http://data.ldeo.columbia.edu

[back to program schedule]

OS62B-0255; EOS, Trans. AGU, 83 (47), F708 (POSTER)

SIOExplorer: Managing Data Flow into a Digital Library

Clark, D; Miller, S P; Peckman, U
dru@sdsc.edu
Geological Data Center, Scripps Institution of Oceanography, UCSD 9500 Gilman Dr., La Jolla, CA 92093-0220 United States
Chase, A; Helly, J
San Diego Supercomputer Center, UCSD 9500 Gilman Dr., La Jolla, CA 92093-0505 United States

The diversity of the data held by the Geological Data Center at SIO is a tribute to the evolution of oceanography, from echo sounding rolls and hand-drawn charts to modern multibeam bathymetry. However, the changes in sensor technology and organizational approaches since 1903 present real challenges to the archivist, as we struggle to migrate the holdings into a web-accessible digital library for use by the public and scientific community (www.nsdl.org). Automation of the dataflow is an absolute necessity, with millions of files and a terabyte of data, although the biggest challenges come from complexity rather than bulk volume. The problems stem from our diverse archive collection, with its evolving data content, processing practices, naming conventions, and clutter from intermediate or obsolete files. This is not the only or the last data system to suffer from these problems, and an approach has been designed that can be applied to other projects. Instead of writing code to process each individual type of cruise, which varies from vessel to vessel over the fifty years and 795 cruises, we created a single Canonical Cruise Data Structure (CCDS). After some experimentation, the CCDS now consists of 9 basic categories and a reasonable number of sub-categories (directories) that can hold all the essential information. The key to flexibility and scalability comes from a template-driven rules approach that allows a processing script to harvest data from complex original data structures, and store them in the simple CCDS. The template mimics the structure of the CCDS. As the processing script traverses the template it finds rules for each category, instructing it on where to look for likely sources, and how to prioritize the results when multiple sources are detected. Over time as new situations are encountered, changes are simply made to the template, rather than the code. After the first tests, it became apparent that a visual method was needed to monitor the success of the harvesting, to make sure that every category is filled with the correct content. We can run simulated tests on our data staging area and report the results immediately in graphical form on our website. The web report shows data found (blue) and not found (red). A mouse-over operation shows the search string used as a rule for selection, and the list of files actually found. We are also employing visualization as a quality control technique to screen data prior to storage in the digital library. We are outputting a grid per multibeam file that can be viewed with public domain GMT tools, the Fledermaus visualization package, or ESRI Arcgis software, as a rapid check on sound velocity artifacts, noise levels, and editing status. http://SIOExplorer.ucsd.edu

[back to program schedule]

OS61C-05; EOS, Trans. AGU, 83 (47), F699

New Magnetic Database Initiatives: Exploitation of and Integration with other Developments

Constable, C; Staudigel, H; Tauxe, L; Koppers, A A P; Johnson, C
cconstable@ucsd.edu
University of California at San Diego, Scripps Institution of Oceanography, La Jolla, CA 92093-0225 United States
Solheid, P; Jackson, M; Banerjee, S
University of Minnesota, Institute for Rock Magnetism , Minneapolis, MN 55455 United States
Pisarevsky, S
University of Western Australia, Tectonics Special Research Center, Crawley, WA 6009 Australia

The steadily increasing collection of paleo, rock, and environmental magnetic data necessitates a community effort to ensure its timely electronic archival and allow appropriate exploitation of research tools from an IT perspective. This will ensure that scientific data gathered with public funds can be readily accessible to the broadest possible range of researchers. Although paleomagnetic databases providing a limited digital archive of legacy data have existed for some time, they lack the interoperability and generality required for the breadth of modern scientific endeavors. Drawing on expertise represented in both the Geochemical Earth Reference Model (GERM) and its parent body EarthRef.org protocols are being developed within the magnetics community that will provide the range of information needed for paleo and rock magnetic databases. The goal in establishing these databases is to provide an enduring digital archive that can be exploited for current scientific investigations, and permit new and interdisciplinary studies that can explore combinations of measurements not previously considered. While strategies for harvesting legacy data are also desirable, a strong focus on gathering newly collected data at the publication stage is necessary, so that these data become broadly available in a timely fashion. The ongoing dialog on minimal and desirable metadata suitable for magnetic databases has a strong basis in the traditional geophysical areas in which magnetic data are

applied, and must also include fundamental rock magnetic information. The metadata must be designed to allow flexible syntheses of magnetic data into the standard kinds of models (such as magnetostratigraphic time scales, geomagnetic field models, plate reconstructions, etc.) and be sufficiently general to enable cross-fertilization with communities involved in related geophysical enterprises such as stratigraphy, petrology, radiometric dating, tectonics and paleoclimate studies. This magnetic database effort lies within the more general hierarchy defined for EarthRef.org, and will exploit both an external and internal modular structure that is intended to facilitate database interoperability. Features that are part of the more general database structure will be illustrated as well as those that are unique to the magnetics database.

[back to program schedule]

OS62B-0258; INVITED; EOS, Trans. AGU, 83 (47), F708 (POSTER)

A metadata scheme for a rock-magnetic data base - considerations and applications

Dekkers, M J; Peters, C
dekkers@geo.uu.nl
Paleomagnetic Laboratory 'Fort Hoofddijk', Utrecht University, Budapestlaan 17, Utrecht, 3584 CD Netherlands
Jackson, M J; Solheid, P
IRM, 291 Shepherd Labs, 100 Union Street SE, Minneapolis, MN 55455-0128 United States

The diverse nature of magnetic property data of rocks slows down the development of a proper data base. This diversity concerns not only the set of parameters determined but also the various types of techniques and instrumentation used, the types of samples, their magnetic and thermal history, and last but not least the diversity of potential users including rock-, paleo- and environmental magnetists, magnetic anomaly modelers, researchers interested in magnetic properties of 'standard' magnetic minerals. Needless to say, these aspects impose special constraints on the data base structure. The metadata scheme proposed begins with the same sample information structure as that of the envisaged paleomagnetic data base, thus optimizing flexibility in interfacing or integrating the databases. For fundamental rock magnetism, various additional fields are required for material properties (including mineralogy, grain size, stoichiometry, defect density, etc.) and experimental conditions (including pretreatments and initial state). The proposed data structure aims to strictly separate the actually measured data and the set of derived parameters. The integrated nature of many magnetic studies will require linking to geochemical and chronostratigraphic data bases in the somewhat further future.

[back to program schedule]

OS62B-0266; INVITED; EOS, Trans. AGU, 83 (47), F710 (POSTER)

Integral Interpretation of Rock Magnetic, Sedimentological and Geochemical Data in Marine Sediment Studies

von Dobeneck, T; Funk, J
dobeneck@uni-bremen.de
University of Bremen Geoscience Dptm., P.O. Box 330440, Bremen, 28334 Germany

Rock magnetic data of marine sedimentary sequences can serve vastly different purposes: (1) identification of NRM carriers, (2) core correlation, (3) orbital age modeling, (4) source and transport tracking, (5) unmixing of terrigenous and biogenous fluxes, and (6) detection of diagenesis effects. Established methods suffice for aim (1) and are often applicable for aims (2-3). For the more ambitious paleoenvironmental aims (4-6), all existing magnetic parameters entail interpretational ambiguities. Magnetic granulometry, for instance, can be an expression of erosion intensity, source mixing, sea level, sorting, winnowing, coarsening by partial depletion or fining by authigenic enrichment. Magnetic concentration parameters and mineral ratios have similar problems. The reliability of paleoclimatic and -oceanographic reconstructions from rock magnetic records depends largely on the correctness of the adapted working hypothesis, which must be newly validated for every location and facies under study. As illustrated by South Atlantic case studies, some ambiguities can be clarified by co-interpreting temporal patterns, regional trends and accumulation rates of selective rock magnetic parameters (e.g. $M_{\{ar\}}$, $M_{\{ir\}}$, $M_{\{hir\}}$ and $K_{\{fd\}}$). For analyzing depositional system, these 'raw' parameters tend to be more valuable than their more frequently published ratios. They can be internally recalibrated to model cumulative parameters such as \kappa or $M_{\{sir\}}$. Conditions for the applicability of this 'partial susceptibilities' method are frequently met in marine settings. An alternative is to calibrate or integrate rock

magnetic with sedimentological and geochemical data. Using porosity, $CaCO_3$, Fe data, an open ocean susceptibility signal can be decomposed into continental (source-mixing), marine (carbonate dilution) and diagenetic (depletion/enrichment) components. The magnetite dissolution index $Fe/K_{\{nd\}}$ quantifies reductive magnetite losses. The carbonate-free dry bulk version of K is reduced to terrigenous influences. To facilitate such analytical methods in future studies, we should not just collect and co-interpret interdiciplinary data sets, but also provide facilities to hold them integrally in rock magnetic data bases.

[back to program schedule]

OS62B-0265; EOS, Trans. AGU, 83 (47), F710 (POSTER)

PMAG: Database Examples for Paleomagnetic and Archeomagnetic Studies

Genevey, A; Tauxe, L
agenevey@ucsd.edu
Scripps Institution of Oceanography, GRD, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0220 United States
Koppers, A A P; Constable, C
Scripps Institution of Oceanography, IGPP, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0225 United States

The Paleo-Magnetic Archival Group (PMAG) designed a modern paleo-magnetic database that is integrated in the existing EarthRef.org umbrella website. PMAG has been developing data and metadata templates that can be used to assemble all data during the publication process in a standardized format. These standardized templates are available in a Microsoft Excel format and can be manipulated using software tools that form the backbone of the data population efforts for this database. In this presentation, we will lay out some examples and show how to use these templates during data population, publication and how they appear in the online databases under http://earthref.org/databases/PMAG/. Using examples for directional paleomagnetic and archeointensity studies, we will explain what data need to be populated, what information is essential or optional, how to customize the data and metadata templates using the template "wizard" (which will hide/show certain tables and columns based on the type of study), how to import/export simple text files, how to validate the data, and how to check for the internal coherence of these data in the template. We will also show how to search online in the EarthRef.org archives and how to download the template files to your own computer. Finally, we will show some basic queries that can be made into (a prototype of) the relational PMAG database, in order to retrieve data. For more information on the development of PMAG data and metadata standards, template files and software tools, please visit the http://earthref.org/metadata/PMAG/ metadata website. http://earthref.org/metadata/PMAG/

[back to program schedule]

OS61C-11; EOS, Trans. AGU, 83 (47), F700

Designing Extensible Data Management for Ocean Observatories, Platforms, and Devices

Graybeal, J; Gomes, K; McCann, M; Schlining, B; Schramm, R; Wilkin, D
graybeal@mbari.org
Monterey Bay Aquarium Research Institute, 7700 Sandholdt Rd., Moss Landing, CA 95039 United States

The Monterey Bay Aquarium Research Institute (MBARI) has been collecting science data for 15 years from all kinds of oceanographic instruments and systems, and is building a next-generation observing system, the MBARI Ocean Observing System (MOOS). To meet the data management requirements of the MOOS, the Institute began developing a flexible, extensible data management solution, the Shore Side Data System (SSDS). This data management system must address a wide variety of oceanographic instruments and data sources, including instruments and platforms of the future. Our data management solution will address all elements of the data management challenge, from ingest (including suitable pre-definition of metadata) through to access and visualization. Key to its success will be ease of use, and automatic incorporation of new data streams and data sets. The data will be of many different forms, and come from many different types of instruments. Instruments will be designed for fixed locations (as with moorings), changing locations (drifters and AUVs), and cruise-based sampling. Data from airplanes, satellites, models, and external archives must also be considered. Providing an architecture which allows data from these varied sources to be automatically archived and processed, yet readily accessed, is only possible with the best practices in metadata definition, software design, and re-use of third-party components. The current status of SSDS development will be presented, including lessons learned from our science users and from previous data management designs.

[back to program schedule]

OS62B-0253; EOS, Trans. AGU, 83 (47), F708 (POSTER)

Scalable Models of Data Sharing in the Earth Sciences

Helly, J J
hellyj@ucsd.edu
San Diego Supercomputer Center, University of California, San Diego 9500 Gilman Drive MC 0527, La Jolla, ca 92093-0505 United States
Staudigel, H; Koppers, A A P
Institute of Geophysics and Planetary Physics, University of California, San Diego 9500 Gilman Drive, la jolla, CA 92093 United States

Many earth science disciplines are currently experiencing the emergence of new ways of data publication and the establishment of an information technology infrastructure for data archiving and exchange. Building on efforts to standardize data and metadata publication in geochemistry, we discuss options for data publication, archiving and exchange. All of these options have to be structured to meet some minimum requirements of scholarly publication, in particular reliability of archival, reproducibility and falsifiability. All data publication and archival methods should strive to produce data bases that are fully interoperable which requires an appropriate data and metadata interchange protocol. To accomplish the latter we propose a new Metadata Interchange Format (.mif) that can be used for more effective sharing of data and metadata across digital libraries, data archives and research projects. This is not a proposal for a particular set of metadata parameters but rather of a methodology that will enable them to be easily developed and interchanged between research organizations. Examples are provided for geochemical and oceanographic data as well as map images to illustrate the flexibility of the approach. http://www.g-cubed.org/

[back to program schedule]

OS62B-0261; EOS, Trans. AGU, 83 (47), F709 (POSTER)

PMAG: Relational Database Definition

Keizer, P; Koppers, A A P; Constable, C; Staudigel, H
pkeizer@ucsd.edu
Scripps Institution of Oceanography, IGPP, University of California, San Diego, LaJolla, CA 92093-0225 United States
Tauxe, L; Genevey, A
Dept. of Geological Sciences, California State Univ., Long Beach, CA 90840 United States
Helly, J
San Diego Supercomputer Center, University of California, San Diego, LaJolla, CA 92093-0527 United States

The Scripps center for Physical and Chemical Earth References (PACER) was established to help create databases for reference data and make them available to the Earth science community. As part of these efforts PACER supports GERM, REM and PMAG and maintains multiple online databases under the http://earthref.org umbrella website. This website has been built on top of a relational database that allows for the archiving and electronic access to a great variety of data types and formats, permitting data queries using a wide range of metadata. These online databases are designed in Oracle 8.1.5 and they are maintained at the San Diego Supercomputer Center. They are directly available via http://earthref.org/databases/. A prototype of the PMAG relational database is now operational within the existing EarthRef.org framework under http://earthref.org/databases/PMAG/. As will be shown in our presentation, the PMAG design focuses around the general workflow that results in the determination of typical paleo-magnetic analyses. This ensures that individual data points can be traced between the actual analysis and the specimen, sample, site, locality and expedition it belongs to. These relations guarantee traceability of the data by distinguishing between original and derived data, where the actual (raw) measurements are performed on the specimen level, and data on the sample level and higher are then derived products in the database. These relations may also serve to recalculate site means when new data becomes available for that locality. The PMAG data records are extensively described in terms of metadata. These metadata are used when scientists search through this online database in order to view and download their needed data. They minimally include method descriptions for field sampling, laboratory techniques and statistical analyses. They also include selection criteria used during the interpretation of the data and, most importantly, critical information about the site location (latitude, longitude, elevation), geography (continent, country, region), geological setting (lithospheric plate or block, tectonic setting), geological age (age range, timescale name, stratigraphic position) and materials (rock type, classification, alteration state). Each data point and method

description is also related to its peer-reviewed reference [citation ID] as archived in the EarthRef Reference Database (ERR). This guarantees direct traceability all the way to its original source, where the user can find the bibliography of each PMAG reference along with every abstract, data table, technical note and/or appendix that are available in digital form and that can be downloaded as PDF/JPEG images and Microsoft Excel/Word data files. This may help scientists and teachers in performing their research since they have easy access to all the scientific data. It also allows for checking potential errors during the digitization process. Please visit the PMAG website at http://earthref.org/PMAG/ for more information. http://earthref.org/PMAG/

[back to program schedule]

OS62B-0244; EOS, Trans. AGU, 83 (47), F706 (POSTER)

Virtual Oregon: A Proof-of-Concept for Seamless Access to Distributed Environmental Information

Keon, D; Pancake, C
keon@nacse.org
Oregon State University, Northwest Alliance for Computational Science and Engineering, Corvallis, OR 97331 United States
Wright, D J
Oregon State University, Department of Geosciences, Corvallis, OR 97331-5506 United States
Walsh, K
Oregon State University, Marine Resource Management Program, Corvallis, OR 97331 United States

Virtual Oregon is a new data coordination center established at Oregon State University in order to: (1) archive environmental and other place-based data on Oregon and associated areas; (2) make those data accessible to a broad spectrum of agencies and individuals via innovative web interfaces; (3) identify key data sets that are not yet available and encourage their collection and dissemination; and (4) facilitate development of statewide standards for archiving, documenting, and disseminating data. Rather than co-locating researchers and data in a physical center, Virtual Oregon employs a distributed architecture that occupies multiple locations while users are presented with the illusion of a single, centralized facility. This approach was selected not just to maximize the impact on campus students, faculty, and staff but also to service broader interactions with extension agents and other members of Oregon State's statewide community. Virtual Oregon builds on regional GIS centers and databanks in a wide range of disciplines, providing decades of research data on topics as varied as coastal processes, climate, biodiversity, land ownership, water quality, wildfire, and agricultural production. There are four distributed nodes, each serving as a center and clearinghouse for distinct types of information and services: - Department of Geosciences (College of Science): geospatial coverages, digital aerial and ortho imagery and associated base data - Forestry Sciences Laboratory (USDA Forest Service and Oregon State's College of Forestry): ecological and resource management databases; data analyses; data from computational simulations - Northwest Alliance for Computational Science and Engineering (NACSE): databases based on specimen collections, field observation, images, or analysis of historical documents; user interface design - Valley Library: published maps, books and archival publications, gray literature, photographs and video Data are harvested from a variety of individuals and research centers and maintained in the distributed nodes using enterprise RDBMS products (Oracle, Sybase, and Microsoft SQL Server) residing on UNIX and Windows platforms. Query Markup Language (QML, a middleware product developed at NACSE) supports database-to-Web interactions by transparently performing queries across multiple RDBMSs and displaying the results as though from a single source. Web-based mapping interfaces (powered by ESRI's Internet Map Server and Spatial Database Engine products) can also be used to explore data visually. In a proof-of-concept under development, users currently have the option of beginning with either the "thematic" or "place-based" interfaces. Ultimately, users will be able to move freely back and forth between the two paradigms, for example initially narrowing the scope of inquiry based on discipline or attributes, moving to the visual interface to refine the search based on location or some set of geospatial characteristics, then moving back to query-based exploration to delve to fine levels of detail. Usability engineering methodologies are being applied so that all navigation and query mechanisms are both maximally productive and easily learned by novices. http://virtual-oregon.nacse.org

[back to program schedule]

OS62B-0263; EOS, Trans. AGU, 83 (47), F709 (POSTER)

The Global Paleomagnetic Database Should Include Specimen-Level Demagnetization Data

Kirschvink, J L
kirschvink@caltech.edu
Division of Geological and Planetary Sciences, California Institute of Technology 170-25, Pasadena, CA 91125 United States
Kazansky, A
Institute of Geology, UIGGM SB Russian Academy of Sciences, Novosibrisk, Sib 000 Russian Federation

In their description of the paleomagnetic database, McElhinny and Smethurst [1] provided a reasonable description of the status quo in terms of compiling and storing a quantitative summary of the results from the past 50 years of paleomagnetic investigation. However, recent developments in science and technology have far outpaced the response of the paleomagnetic community on the issue of data archive and storage, and as a consequence this field risks being left in the dust. It is our opinion that the database should be expanded to include the entire set of specimen-level demagnetization data that go into a paleomagnetic study, along with the interpretation made by the authors. A simple back-of-the-envelope estimate suggests that all the raw paleomagnetic demagnetization data generated in the past 50 years would fit easily on a single hard drive of a personal computer. Other fields are doing precisely this type of the data archiving -- for example, nothing is accepted for publication in the good, peer-reviewed molecular genetics journals until the raw sequence data have been deposited in one of the internationally-accepted genetic databanks. As we feel that geophysics is in no way inferior to genetics, we should have no fear of archiving our data and making it publicly available in a similar fashion. The actual process of measuring and demagnetizing typical paleomagnetic materials has not changed much in the 30 years since the introduction of superconducting magnetometers, with most studies still utilizing some combination of alternating field, thermal, microwave, or chemical demagnetization to isolate the principal magnetic components (although the number of demagnetization steps per specimen has certainly improved with the increasing use of automation). In contrast, techniques for the analysis of paleomagnetic data have improved substantially (e.g., [2,3]). Hence, public access to the archived raw demagnetization data would not only allow independent scrutiny of the interpretive stage of paleomagnetic data analysis, it would permit new techniques with potentially greater analytical ability to be applied as they are developed for the recognition of otherwise hidden magnetic components in previously published studies. Most of the popular routines for performing principal component analysis on specimen-level demagnetization data readily allow the importation of data into their fixed format files structure, and the data can be archived and indexed easily in this fashion (see [4,5]). We suggest that the deposition of such data in an expanded paleomagnetic repository be included as an eighth criteria in the commonly-used 7-point Van der Voo reliability scale, and that deposition such data be required before publication in any AGU-sponsored journal.

[back to program schedule]

OS61C-13; EOS, Trans. AGU, 83 (47), F700

The CompreHensive collaborativE Framework (CHEF)

Knoop, P A
knoop@umich.edu
University of Michigan, School of Information 1075 Beal Ave., Ann Arbor, MI 48109-2112 United States
Hardin, J
University of Michigan, Media Union 2281 Bonisteel Blvd., Ann Arbor, MI 48109-2094 United States
Killeen, T; Middleton, D
National Center for Atmospheric Research, 1850 Table Mesa Dr., Boulder, CO 80305 United States

Data integration, publication, and archiving have become important considerations in most fields of science as experiments and models increase in complexity, and the collaborations necessary to conduct the research grow broader. The development of well thought out strategies and standards for such data handling, however, only goes part way in supporting the scientific process. A primary driving force for such efforts is the need of scientists to access and work with data in a timely, reasonable, and often collaborative fashion. Internet-based collaborative environments are one way to help complete this picture, linking scientists to the data they seek and to one another (e.g., Towards a Robust, Agile, and Comprehensive Information Infrastructure for the Geosciences: A Strategic Plan For High Performance Simulation, NCAR, 2000, http://www.ncar.ucar.edu/Director/plan.pdf). The CompreHensive collaborativE Framework (CHEF, http://chefproject.org) is a generic, extensible, web-based, open-source environment for collaboration. CHEF's goal is to provide the basic building blocks from which a community can assemble a collaborative environment that fits their needs. The design of CHEF has been influenced by our experience developing the Space Physics and Aeronomy Research Collaboratory (SPARC, http://www.si.umich.edu/SPARC), which provides integrated access to a wide variety of heterogeneous data sources, including community-standardized data bases. The design has also been heavily influenced by our involvement with an effort to extract and codify

the broad underlying technical and social elements that lead to successful collaboratories (http://www.scienceofcollaboratories.org). A collaborative environment is in itself also not the complete answer to data handling, rather, it provides a facilitating environment in which community efforts to integrate, publish, archive, and share data using standard formats and practices can be taken advantage of by the end-users, the scientists. We present examples of how CHEF and its predecessors are utilized in a wide variety of scientific communities, including engineering, chemistry, and the geosciences. In particular, we focus on CHEF's utilization by the earthquake engineering community, whose Network for Earthquake Engineering Simulation (NEES, http://www.nees.org) involves a community effort to develop data standards and practices. In this context NEES is using CHEF as the "integration" environment in which to place the "tools" that bring together scientists and data; this includes data browsers, meta-data search engines, real-time and archival data viewers, etc. By developing these tools within the CHEF framework and exposing the community-developed data standards to the framework, they automatically gain the features, functionality, and capabilities offered by the collaborative environment. We also explore how a collaborative environment, in conjunction with community developed standards and practices for data integration, publishing, and archiving, could benefit the ocean science community. http://chefproject.org

[back to program schedule]

OS62B-0260; EOS, Trans. AGU, 83 (47), F709 (POSTER)

PMAG: Database Development Under the EarthRef.org Umbrella Website

Koppers, A A P; Constable, C; Staudigel, H
akoppers@ucsd.edu
Scripps Institution of Oceanography, IGPP, University of California, San Diego, LaJolla, CA 92093-0225 United States
Tauxe, L
Scripps Institution of Oceanography, GRD, University of California, San Diego, LaJolla, CA 92093-0220 United States
Helly, J
San Diego Supercomputer Center, University of California, San Diego, LaJolla, CA 92093-0527 United States

EarthRef.org has been compiling resources used to construct Earth Reference Models such as for geochemistry (GERM) and geophysics (REM). This approach can readily be expanded to any type of Earth science data and, to that extent, we have started an online database for paleomagnetism (PMAG) under http://earthref.org/databases/PMAG/. This database stores all measurements and their derived properties for studies on paleomagnetic directions (inclination, declination) and intensities. Ultimately, this database will allow researchers to study on the internet and download important data sets that, for example, display the paleo-secular variations in the intensity of the Earth's magnetic field over geological time. The PMAG database is completely integrated in the EarthRef.org relational database structure and thus benefits significantly from already-existing common database components, such as the EarthRef Reference Database (ERR) and Address Book (ERAB). The ERR allows researchers to find complete sets of literature resources as used in either GERM, REM or PMAG. The ERAB contains the addresses for all contributors to the EarthRef.org databases, but also for all the rock collectors, archivers and analysts appearing in these databases. Integration with both components will ensure direct traceability to the original sources of the PMAG data and metadata. Data contribution to the PMAG database is most critical in achieving an useful research tool. We have been developing data and metadata templates that can be used to provide all data during the publication process in a standardized format. Software tools are provided to facilitate an easy population of these templates. These tools allow for the import/export of data files in a delimited text format, and they provide functionality to validate data and to check their internal coherence in the template. During and after publication these standardized PMAG templates will be stored in the ERR database. From that moment on they can be searched for and downloaded from the EarthRef.org website. Finally, the contents of these template files will be automatically read out and parsed into the online relational PMAG database. For more information on the development of PMAG data and metadata standards, template files and software tools, please visit the http://earthref.org/metadata/PMAG/ metadata website. http://earthref.org/metadata/PMAG/

[back to program schedule]

OS62B-0245; EOS, Trans. AGU, 83 (47), F706 (POSTER)

New Challenges in Sample-Based Data Implementation

Lehnert, K A
lehnert@ldeo.columbia.edu
Lamont-Doherty Earth Observatory, Columbia University, 61 Route 9W, Palisades, NY 10964 United States
Lenhardt, C
Center for International Earth Science Network CIESIN, Columbia University, 61 Route 9W, Palisades, NY 10964 United States
Langmuir, C H
Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, MA 02138 United States

Geochemical data on rocks are now widely available in relational databases (PetDB, GEOROC, NAVDAT) served over the internet, using the schema of Lehnert et al. (2000). The widespread access to data has significantly increased the overall efficiency of researchers. It has aided the complete range of research activities including teaching, proposal writing and publication, and allows people from related fields to address geochemical problems (e.g Kellogg 2002). Future advances have to include more than simply adding more analytical data in the same formats: (1) A scheme of universal sample identification will be essential for terrestrial samples to allow compilation of data from different sources on the same sample. How to develop and implement such a scheme is challenging and should happen in coordination and cooperation with existing and emerging efforts for terrestrial sample archives (e.g. Goldstein et al., 2001). (2) The ability to integrate geochemical with geophysical and geological data is vital. Visualization of data on maps and comparison of geochemical data to grid-based data presentations requires effective integration of relational databases with GIS and other information technologies such as XML. The industrial and government orientation of widespread commercial products may not satisfy the specific and flexible needs of researchers. (3) Easy access to other published material needs to be dealt with in terms of intellectual property. For example, published sample location maps are often not available in digital format, and reproduction of those (copyrighted) maps from publications, with links to the data would be valuable. It would also be useful to provide access to pdf files of complete publications through a simple database link, which would require a different level of cooperation with publishers. (4) Derivative products that compile data from different sources and their presentation on appropriate maps or other visualizations would be a further advantage for a wide variety of research problems. Automated update of such products would require successful resolution of the previous three challenges. Goldstein, S.L., W. Melson, Geochemical News: 108: 19-20, 2001. Kellogg, J.B. et al., Geochim Cosmochim Acta 66, A391, 2002. Lehnert, K. et al., Geochemistry, Geophysics, Geosystems 1, 2000. http://www.ldeo.columbia.edu/RidgePetDB

[back to program schedule]

OS62B-0249; EOS, Trans. AGU, 83 (47), F707 (POSTER)

Using GeoVRML for Visual Dissemination of Oceanographic Data

McCann, M P
mccann@mbari.org
Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA 95039-9644 United States

Visual representation of three dimensional geospatial information is an often requested feature of oceanographic data management systems. The Monterey Bay Aquarium Research Institute's Expedition Database provides access to data from over 2500 submersible dives beginning in 1988. Visualizations of these dives have been produced and are now accessible through a simple web browser interface. The World Wide Web, database management software, and 3D graphics processing units have advanced to enable simplified viewing of complex data. However; issues such as single precision graphics pipe-line arithmetic and geographic coordinate transformations complicate the presentation of geospatial data. GeoVRML is an open international standard that has been developed to address these issues. In addition to being an ISO standard, GeoVRML provides tools and recommended practices for representing 3D geographic data. Any sort of geospatial data can be represented in GeoVRML including high-resolution bathymetric data, submersible dive tracks, remotely sensed imagery, and animated 3D objects. Furthermore, interactivity with the data may be provided with standard Virtual Reality Modeling Language (VRML) scripts and prototypes. Content placed in GeoVRML format is viewable inside a web browser and can be integrated with web-based data delivery systems to provide easily understood visual representations of geospatial data. MBARI uses GeoVRML to disseminate 3D replays of submersible dive data stored in its Expedition Database. Terrain data, ship and vehicle navigation, environmental

(CTDO) data, video frame grabs, samples data, and video annotation information can all be viewed together using this tool. Tools to generate GeoVRML terrain content have been developed and are provided in the open-source tsmApi and MB-System packages. These tools may be used to convert bathymetric data into multi-resolution quad-tree hierarchical tiles that load efficiently over wide area networks. The level of detail for the terrain is controlled by the user who may interactivly select terrain tiles for higher resolution display. Up to 9 levels of detail (10 m resolution over a distance of 55 km) are effectively visualized using this method. http://www.mbari.org/~mccann/vrml/ROVDataVis

[back to program schedule]

OS61C-06; EOS, Trans. AGU, 83 (47), F699

SIOExplorer: Overview, Initial Results and Next Steps

Miller, S P; Peckman, U; Clark, D
spmiller@ucsd.edu
Geological Data Center, Scripps Institution of Oceanography, UCSD 9500 Gilman Dr., La Jolla, CA 92093-0220 United States
Helly, J J
San Diego Supercomputer Center, UCSD 9500 Gilman Dr., La Jolla, CA 92093-0505 United States
Africa, M
Major, Hagen and Africa, 500 Washington St., San Francisco, CA 94111 United States
Day, D
UCSD Libraries, UCSD 9500 Gilman Dr., La Jolla, CA 92093-0219 United States

Data, documents and images from 795 expeditions by the Scripps Institution of Oceanography (SIO) since 1903 are becoming web-accessible for both education and research through the new SIOExplorer project (http://SIOExplorer.ucsd.edu), which is a collection in the overall NSF-funded National Science Digital Library (www.nsdl.org). The collaborative effort includes researchers at SIO, computer scientists from the San Diego Supercomputer Center (SDSC), and archivists and librarians from the UCSD Library. The co-authors of this paper tested a shipboard prototype during a Floating Digital Library Workshop from New Zealand to Samoa on R/V Melville in March, 2002. General purpose tools have been developed to automate collection development, manage metadata, and geographically search the library, as discussed in other presentations in this session. In the initial year of operation, the biggest challenge has been wrestling with the volume and variability of data and documents. Shipboard sensors, data volumes, and organizational structures have evolved greatly over the decades, particularly with 244 multibeam expeditions since 1982. Considerable success came after introducing the concept of a Canonical Cruise Data Structure (CCDS) with nine basic categories that seem to capture the essential characteristics of data practices since the 1960's. Automatic software pulls data into the CCDS from diverse source directories and media, guided by a template with rules for priority and filenames. Almost all metadata are harvested automatically into simple "metadata interchange format" (.mif) files, one for each "arbitrary digital object" (ADO) in the CCDS. The metadata are placed in an Oracle database, and the associated data are managed by the SDSC Storage Resource Broker on various disk and automatic tape silo systems. The system is extensible to various domains and data types, including geochemistry, image archives, multibeam bathymetry, reports and publications. A Java Metadata Object Browser and Editor (MOBE) expands or hides the complexity for each domain, as needed. A prototype interactive CruiseViewer with both Java and html approaches will be demonstrated. As the second year of the project begins, greater emphasis will be placed on search and display tools. At-risk data on shipboard magnetic tapes will be migrated to RAID systems and tape silos. Public outreach will begin at the Birch Aquarium and other locations. A workshop will be held at Scripps in September 2003, coinciding with the hosting of the Oceans 2003 meeting and the 100th Anniversary of SIO. These efforts are supported by the NSF NSDL and ITR programs and by SIO institutional funds. http://SIOExplorer.ucsd.edu

[back to program schedule]

OS61C-08; EOS, Trans. AGU, 83 (47), F700

The Evolution of Global Oceanic Crust From Jurassic to Present Day: A Global Data Integration

Mueller, R D; Gaina, C; Clark, S; Sdrolias, M
dietmar@es.usyd.edu.au
School of Geosciences, Edgeworth David Building F05 University of Sydney, Sydney, NSW 2006 Australia
Roest, W R
Geological Survey of Canada Natural Resources Canada, 615 Booth Street, Ottawa, ON K1A 0E9 Canada

Understanding the interplay between oceanic crustal production, mid-ocean ridge volumes, mantle convection, and long-term sea level changes is one of the first-order goals of Earth science. In order to investigate whether the oceanic crustal production varied through time we created complete digital paleo-seafloor age grids for the last 180 million years. A combination of methods and datasets were used to map and reconstruct preserved and subducted oceanic regions. Magnetic anomalies and gravity anomaly derived from satellite altimetry were used to derive the age of oceanic crust, spreading rates, geometry and grain of seafloor. We have created an extensive GMT database for the magnetic and gravity data and have also used the magnetic database of GSC for North Atlantic magnetic data. In addition we have used the predicted bathymetry and the ETOPO5 bathymetry grids for mapping various tectonic features of the ocean floor. Available seismic data have been also used to constrain the geometry of some parts of the oceanic basins. A compilation of geological data from published studies (type and ages of dredged rocks, evidence of subduction related magmatism and metamorphism, ophiolites) in conjunction with published results from ODP and DSDP cruises were employed to constrain and groundtruth our models. A subduction related set of geological data is partially compiled in a database(Microsoft Access). The tectonic history of various oceanic areas has been revisited using both quantitative and qualitative methods. We have used a quantitative method based on Hellinger's criteria of fit to derive the evolution of North Atlantic and other small basins east of Australia (Tasman and Coral seas). For regions with sparse magnetic data coverage or complicated seafloor spreading pattern we have used qualitative methods, that usually involve visual fit reconstructions (PLATES software). This method has been used to derive the evolution of a series of backarc basins in the SW Pacific and SE Asia and in the Indian Ocean. The new results have been integrated in a global tectonic model and newly constructed isochrons were used to update the present day oceanic agegrid. Paleo-oceans are modelled by creating synthetic plates whose locations and geometry is established on the basis of preserved M-sequence magnetic lineations, paleogeography, regional geological data and the rules of plate tectonics. Plate boundaries, which are introduced and modified in time and space, give only little room for alternative plate model solutions. These plate limits are governed by rheological laws, which provide stable constraints for reconstructions when geological information is scarce. This method has been used to reconstruct subducted Neo-Tehys ocean and the Izanagi/Kula, Farallon and Phoenix plates. The relative plate motion models used to derive isochrons were linked to an absolute plate model to reconstruct isochrons in a desired framework, in order to produce gridded paleo-age maps of the ocean floor that will provide an accurate estimation of seafloor spreading/subduction rates, as well as the global distribution of oceanic crust ages for the last 180 million years. The paleo-age grids illustrate where subduction zones were located in time and provide constraints for geodynamic models, for the heat loss of the Earth, and for estimates of plate driving forces through time.

[back to program schedule]

OS62B-0262; EOS, Trans. AGU, 83 (47), F709 (POSTER)

Paleointensity Database: Current State and Future

Perrin, M
perrin@dstu.univ-montp2.fr
Laboratoire GTS UMR CNRS-UM2 5573, Universite Montpellier II CC60, Montpellier Cedex 5, 34090 France

IAGA Working Group I-3 (Paleomagnetism) and I-4 (Rock Magnetism) initiated the establishment of various databases to cover most types of paleomagnetic data, among which a paleointensity database. This database comprises only absolute paleointensity determinations from igneous rocks and baked contacts, archaeological artifacts are not included. All methods of paleointensity determinations are taken into account as well as all field configurations (normal, reverse or transitional polarity). Around 2.500 data are now available from almost 200 references. For practical reasons, the paleointensity database was distinct from the global paleomagnetic database. Also only mean results for given cooling units were registered in the paleointensity database. For the future, it will be desirable to construct a more

extensive database which will incorporate not only mean estimates per cooling but also the raw data at the specimens level and all necessary metadata as rock magnetism, geology, petrology, radiometric datings . Paleointensity data should also be directly linked to global paleodirectional data, as well as with archaeological determinations. This will allow an easier analysis of the total magnetic field and a better way to assess the reliability of the determinations. In some cases, it will also allow reanalysis of the raw data, following improvements in the theory of magnetic acquisition and/ or updating of the determination methods.

[back to program schedule]

OS62B-0259; INVITED; EOS, Trans. AGU, 83 (47), F709 (POSTER)

Visual Palaeomagnetic Database

Pisarevsky, S
spisarev@tsrc.uwa.edu.au
Tectonics Special Research Centre, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009 Australia

The Global Palaeomagnetic Database (GPMDB) created and developed by McElhinny and Lock (1991, 1996) is used by researchers all over the world. The user-friendly interface makes it unnecessary for users to learn about details of the Microsoft Access software. The next step in the development of these databases lies in the visualisation of data and in the integration of the palaeomagnetic data with Geographical Information Systems (GIS). One of the most popular GIS software among Earth scientists is ArcView. Due to the relative simplicity of the structure of the GPMDB it quite easy to integrate palaeomagnetic data into GIS. It is just necessary to prepare a database file in DBASE format (using the export option of the Microsoft Access, for example) and then to create a subsequent graphic theme (layer) in ArcView. The wide variety of ArcView options enable the use of graduated colours, labels, and different symbols to emphasise ages, palaeomagnetic directions, or other data features. Palaeomagnetic data may be combined with the geological, tectonic, and other maps using a variety of spherical projections. Palaeomagnetic data may be easily integrated into other GIS-oriented databases, such as geochronological databases. In addition to all "traditional" services known for the GPMDB users, such as queries, ArcView and supplementary Avenue scripts provide many new possibilities. For example, it is very easy now to choose data from a particular polygon (e.g. craton, terrane, orogenic belt etc.). Users of the new visual database also can "instantly" create a stereoplot for any selected data subset and to calculate mean directions and palaeopoles. It is also possible to display palaeopoles for the selected group of data and to reconstruct a palaeoposition of the continental block using these poles, or Euler pole of rotation. There are obvious advantages of using the visual database. For example, if there are some errors in the geographical position of some data, in many cases it is easy to find them. It is much easier now to test tectonic hypotheses. For instance, it is possible to compare data from two adjacent terranes to decide the time of their collision.

[back to program schedule]

OS62B-0248; EOS, Trans. AGU, 83 (47), F707 (POSTER)

Earth Science Markup Language: An Overview

Ramachandran, R; Conover, H; Graves, S
rramachandran@itsc.uah.edu
Information Technology and Systems Center, University of Alabama in Huntsville, S339 Technology Hall, Huntsville, AL 35899 United States

The Earth Science Mark-up Language (ESML) is an effort funded by the NASA Earth Science Technology Office and designed by the Information Technology and Systems Center (ITSC) at the University of Alabama in Huntsville (UAH) as a new approach for allowing data/application interoperability. This new descriptive language, based on the eXtensible Markup Language (XML) provides a structured way to describe not only the content and structure of a data file or collection, but also the semantic information. This combination of structural and semantic information allows an application to intelligently interpret the data, thus facilitating development of search, visualization, and analysis tools that are independent of data type or format. By defining a standard for external metadata to describe the content, structure, and semantics of a file, ESML provides a means for applications to utilize legacy, current, and future data sets in an integrated fashion. This paper will provide an overview of the ESML project, its goals, how various data users and application developers could benefit from ESML, and some selected applications that are utilizing this technology. http://esml.itsc.uah.edu

[back to program schedule]

OS62B-0267; EOS, Trans. AGU, 83 (47), F710 (POSTER)

Performance Evaluation of INMARSAT Fleet 77 Services Aboard the R/V Ewing

Schmidt, V E; Chayes, D N; Gold, E
vschmidt@ldeo.columbia.edu
Lamont-Doherty Earth Observatory of Columbia University, 65 Route 9W, Palisades, NY 10964 United States

In late 2001, the R/V Ewing was asked to conduct a trial installation of the Thrane and Thrane [of Denmark] F77 antennae [TT-3084A Capsat] and the newest IMNARSAT communications channel, INMARSAT F. Branded as "Fleet F77 Service" by INMARSAT, the service provides ISDN 64kbps and 56kbps high quality voice and data connections as well as Mobile Packet Data Service which allows an "always on" connection under which users pay for the packets they send rather than the time they are connected. Fleet F77 also allows low bandwidth [2.4kbps] "Mini-M" voice and fax services. While not currently available, the Capsat antennae also is prepared to take advantage of 4th Generation Inm-IV satellites [expected in 2004] allowing LAN speeds up to 432kbps. The F77 antenna consists of two units, the TT-3084A antenna and a single "Below Deck Unit". The Capsat antennae radome is a mere 84 cm in diameter - considerably smaller than that typically associated with INMARSAT A or B. It was mounted above the forward port corner of the pilot house atop a reinforced mast. Below deck electronics consist of a single unit containing three analog RJ-11 interfaces, a single ISDN interface, two RS-232 serial interfaces, a USB interface [not functional on our test unit] and a standard Handset. This was mounted in the pilot house electronics space. The Capsat antennae and associated electronics were installed in Guam in mid February 2002 and the system began operational trials during the following cruise on February 24th. Tests of the Fleet 77 system consisted of Mini-M voice and fax both to and from the ship, 64kbs voice to and from the ship, MPDS connects to shore, and operational tests with the INMARSAT Command Center. The trial period completed May 12th after which the F77 became an integral part of the Ewing's communication suite. Results of these tests as well as latency and packet loss measurements made over various data connection types will be presented.

[back to program schedule]

OS62B-0257; EOS, Trans. AGU, 83 (47), F708 (POSTER)

Potential of Scalable Vector Graphics (SVG) for Ocean Science Research

Sears, J R
jsears@agu.org
American Geophysical Union, 2000 Florida Avenue NW, Washington, DC 20009 United States

Scalable Vector Graphics (SVG), a graphic format encoded in Extensible Markup Language (XML), is a recent W3C standard. SVG is text-based and platform-neutral, allowing interoperability and a rich array of features that offer significant promise for the presentation and publication of ocean and earth science research. This presentation (a) provides a brief introduction to SVG with real-world examples; (b) reviews browsers, editors, and other SVG tools; and (c) talks about some of the more powerful capabilities of SVG that might be important for ocean and earth science data presentation, such as searchability, animation and scripting, interactivity, accessibility, dynamic SVG, layers, scalability, SVG Text, SVG Audio, server-side SVG, and embedding metadata and data. A list of useful SVG resources is also given.

[back to program schedule]

OS61C-01; INVITED; EOS, Trans. AGU, 83 (47), F698

Data Integration Across the Geoscience Disciplines: Challenges and Opportunities

Snyder, W S
wsnyder@nsf.gov
Earth Sciences Division National Science Division, 4201 Wilson Blvd Rm 785, Alrington, VA 22230 United States

As geoscience research becomes more interdisciplinary and integrative, it is also becoming increasingly dependent on rapid access to reliable information. Some of these data sets are extremely large and new data sets are being developed. Interoperability among data sets is of paramount importance and researchers are pushing for the development of new analytical tools. To better understand the full range of earth processes, community models and new theoretical frameworks are being developed that require increased computational capabilities. Within the NSF Geoscience Directorate, the three divisions (Earth, Atmospheric, and Ocean sciences) are working with their communities to formulate informatics programs. Whereas each division serves distinct communities, the various subdisciplines have overlapping informatics needs that require a mechanism to serve both their specific requirements as well as one that promotes syngerism among the sciences. For the Earth Sciences community, there is a natural division into at least four earth science-based informatics groups: 1) solid

earth geophysics and active tectonics; 2) continental crustal evolution and architecture; 3) surficial processes and hydrologic sciences; and 4) sedimentary and ancient life systems. In addition, education-outreach and computational technology are critical parts of the overall system. Each of these components encompasses several ongoing or developing informatics efforts, such as those by IRIS, EarthScope, several NSF Information Technology Research (ITR) grants, GERM, NAVDAT, the Hydrologic Information System (CUAHSI), the Community Sediment Model (CSM), CHRONOS, and many others. GEON (A Research Project to Create Cyberinfrastructure for the Geosciences) is an ITR NSF project that perhaps forms a core part of the computational facilities for the earth sciences and it includes some science-based projects that are encompassed in the respective discipline based groupings. Neither the names or the "membership" in these topical groups are firmly establish. Furthermore, because informatics must reflect and serve the community needs, everyone who has an interest in or need for informatics must be provided an opportunity to become part of the effort. On the other hand, we must have a limited number of science-based categories or the funding and coordination of efforts becomes untenable. It is important to emphasize that there are clear overlaps between these earth science-based efforts and similar to parallel ones in the ocean and atmospheric sciences. The PETDB and ODP's Janus databases are two such academic-based examples that connect the earth science and oceanographic communities. Atmospheric and hydrologic scientists are working to bridge their information systems. Similarly, a need to cross the interface between the geosciences and the ecosystem and modern life sciences is being articulated. Representatives from federal and state agencies and industry sit on many of subdiscipline organizing and steering committees. The important point is that the scientists are the ones articulating the need for this informatics integration, and therefore it appears that informatics is becoming a bottom-up driver for better overall science integration.

[back to program schedule]

OS62B-0264; EOS, Trans. AGU, 83 (47), F710 (POSTER)

Archival and Retrieval of Multi-Dimensional Rock Magnetic Data: A Job for BLOB's

Solheid, P A; Jackson, M J; Marvin, J A; Banerjee, S K
peat@umn.edu
Institute for Rock Magnetism, 291 Shepherd Labs 100 Union St. SE, Minneapolis, MN 55455 United States
Kamp, W P
IAG Partners, Inc., 12904 Hamlet Ave, Apple Valley, MN 55124 United States

Rock magnetic data range in complexity from simple parameters such as SIRM, to structured multidimensional data sets such as First Order Reversal Curves (FORC's) and magnetic susceptibility measured as a function of temperature, amplitude, and frequency. While some data are suited to storage in a conventional database, such as age- or depth-varying parameters common to environmental magnetism, more complex multi-dimensional data sets can not easily be stored in a database. However, these complex data sets can be stored as BLOB's (Binary Large Objects), in separate data files with catalogues and derived parameters in the actual database. Programs or applets can then be written to search through all or a subset of BLOB's for certain information or characteristics that are stored in the data sets. However, searching through a large number of BLOB's is a slow process. It is necessary to provide a fairly comprehensive summary in the forms of metadata, summary data and derived data, which are stored in the database and can be used to limit the scope of the BLOB search. We will illustrate several sets of data from simple to complex and show how the metadata, summary data and derived data relate to the various primary data. We will also propose some initial methods or applets that can be used to search BLOB's and extract useful information.

[back to program schedule]

OS61C-04; EOS, Trans. AGU, 83 (47), F699

GERM in EarthRef.org: A reference model approach to data bases

Staudigel, H; Koppers, A A P; Constable, C
hstaudigel@ucsd.edu
Scripps Institution of Oceanography, UCSD-0225, La Jolla, CA 92093-0225 United States
Helly, J
San Diego Supercomputing Center, UCSD, La Jolla, Ca 92093 United States

The geochemical earth reference model (GERM) initiative is a grass-roots effort that works towards a chemical characterization of the whole earth, from the core to the atmosphere. This reference model aims to provide the currently best possible geochemical constraints that characterize every geochemical reservoir, as a reference point for comparison with data from future studies to improve these estimates. The GERM reference model approach defines a scientific process with a well structured, overarching scientific goal in geochemistry that is central to how geochemistry works and interacts with neighboring disciplines. This goal of a reference model prescribes a clear path that helps optimize a data model and a data structure that makes it easy to find, use and compare data. Earthref.org has designed a hierarchical

metadata scheme and developed a data base structure for GERM that is modular and optimized for the GERM process. We recently began applying these techniques to other data bases such as in paleomagnetism and it is quite obvious that same approach can be rather widely applied. Most geoscience disciplines have central reference models that define a discipline in its scientific approach and data base needs. Such models may include e.g., bio- and magnetostratigraphy, plate motion, or the physical structure of the earth. The contents and structure of these reference models may vary dramatically between earth science disciplines but they all have in common that they offer basis for a meaningful information technology infrastructure that is transparent and tailored to specific science goals, and makes data accessible to neighboring disciplines In our presentation, we will summarize the structure and contents of GERM data base and the learning path it took to realize this data base. There are many lessons learned and many concepts developed that are useful in the development of other data bases to serve their community. Key aspects of this GERM learning process included: - Community involvement is essential to define the scientific focus of data base development, to establish widely accepted data and metadata formats, to enable and encourage the community to use and contribute online tools and database contents. - Data base efforts have to interface with data publication in order to allow for publication of important data that would be otherwise not published and to shift data publication methods away from traditional printed data tables towards database ready data and metadata files. - Metadata and data and their archiving infrastructure have to be modular, transparent and maximally transportable between data bases. - Databases have to involve the community in the maintenance of legacy data, through the encouragement of publication and through uploading features in the data base. The EarthRef.org database architecture and metadata structure has a very modular character and is organized in a hierarchical fashion, such that they can be easily described in a canonical database structure and easily transported to other disciplines. EarthRef.org has used these structures to jump-start a database initiative in palaeomagnetism.

[back to program schedule]

OS62B-0254; EOS, Trans. AGU, 83 (47), F708 (POSTER)

CruiseViewer: SIOExplorer Graphical Interface to Metadata and Archives

Sutton, D W; Helly, J J; Chase, A
suttond@sdsc.edu
San Diego Supercomputer Center, 9500 Gilman Dr, La Jolla, CA 92093-0527 United States
Miller, S P; Clark, D
Geological Data Center, Scripps Institution of Oceanography, 9500 Gilman Dr, La Jolla, CA 92093-0220 United States

We are introducing "CruiseViewer" as a prototype graphical interface for the SIOExplorer digital library project, part of the overall NSF National Science Digital Library (NSDL) effort. When complete, CruiseViewer will provide access to nearly 800 cruises, as well as 100 years of documents and images from the archives of the Scripps Institution of Oceanography (SIO). The project emphasizes data object accessibility, a rich metadata format, efficient uploading methods and interoperability with other digital libraries. The primary function of CruiseViewer is to provide a human interface to the metadata database and to storage systems filled with archival data. The system schema is based on the concept of an "arbitrary digital object" (ADO). Arbitrary in that if the object can be stored on a computer system then SIOExplore can manage it. Common examples are a multibeam swath bathymetry file, a .pdf cruise report, or a tar file containing all the processing scripts used on a cruise. We require a metadata file for every ADO in an ascii "metadata interchange format" (MIF), which has proven to be highly useful for operability and extensibility. Bulk ADO storage is managed using the Storage Resource Broker, SRB, data handling middleware developed at the San Diego Supercomputer Center that centralizes management and access to distributed storage devices. MIF metadata are harvested from several sources and housed in a relational (Oracle) database. For CruiseViewer, cgi scripts resident on an Apache server are the primary communication and service request handling tools. Along with the CruiseViewer java application, users can query, access and download objects via a separate method that operates through standard web browsers, http://sioexplorer.ucsd.edu. Both provide the functionability to query and view object metadata, and select and download ADOs. For the CruiseViewer application Java 2D is used to add a geo-referencing feature that allows users to select basemap images and have vector shapes representing query results mapped over the basemap in the image panel. The two methods together address a wide range of user access needs and will allow for widespread use of SIOExplorer. http://sioexplorer.ucsd.edu

[back to program schedule]

OS62B-0247; EOS, Trans. AGU, 83 (47), F707 (POSTER)

Implementation and compatibility of a North American Volcanic and Plutonic rock database (NAVDAT)

Walker, J D; Bowers, T D; Black, R A
jdwalker@ku.edu
Department of Geology, University of Kansas, Lawrence, KS 66045 United States
Glazner, A F; Straus, A K
Department of Geological Sciences, University of North Carolina, Chapel Hill, NC 27599 United States
Farmer, G L
Department of Geological Sciences, University of Colorado, Boulder, CO 80309 United States
Carlson, R W
Carnegie Institution of Washington, 5241 Broad Branch Rd, NW, Washington, DC 20015 United States
Grossman, J N
US Geological Survey, 954 National Center, Reston, VA 20192 United States

NAVDAT is a database for igneous rocks in western North America that will contain geochemical and age information primarily on Cenozoic rocks. NAVDAT will allow exploration of temporal and spatial patterns in igneous activity, and to connect these patterns with local and regional tectonic development and lithospheric structure. The database will be web-accessible for downloads and queries (navdat.geo.ku.edu). Allied information, such as geologic and geophysical maps, crustal structure, etc., will also be available through a map interface. We have attempted to keep the NAVDAT schema compatible with that for PetDB and GEOROC (petdb.ldeo.columbia.edu, georoc.mpch-mainz.gwdg.de) in order to build consensus on an overall structure for of an igneous rock database. The issues to be addressed by continent-based NAVDAT, however, are somewhat different from ocean-floor based PetDB, and the schema required numerous modifications. We have extended the schema in several areas to meet the needs of the on-land database. Location and age information become critical because we are trying define changes in magma source with time tied to structural position and setting. For this reason, we have added more fields to cover such issues as how a rock is dated and where it is located. In addition, we have implemented an expanded reference section that imports all information available in AGI's Georef database. This should allow for superior query ability. One recurring issue in constructing the NAVDAT database is the inconsistency in the way geochemical data are reported. The following is a suggested publication check-list for geochemists that will enable more robust database construction: 1) all samples must have locations reported as accurately as possible, not just located on a map figure or given as a general location; 2) known sample ages must be given and the method of dating explained (e.g., directly dated, stratigraphically bracketed, or correlated in a regional sense); 3) laboratory techniques must be documented, and include both where and how analyses were done; 4) it must be made clear when reporting data for a standard whether the other data presented have been renormalized using an accepted value for the standard (especially for isotopic data).

[back to program schedule]

OS62B-0250; EOS, Trans. AGU, 83 (47), F707 (POSTER)

Video data annotation, archiving, and access

Wilkin, D; Connor, J; Stout, N J; Walz, K; Schlining, K; Graybeal, J
wilkin@mbari.org
Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road , Moss Landing, CA 95039 United States

Scientifically useful, high-quality video data can be challenging to integrate with other data, and to analyze and archive for use in ocean science. The Monterey Bay Aquarium Research Institute (MBARI) uses high-resolution video equipment to record over 300 remotely operated vehicle dives per year. Over the past 14 years, 13,000 videotapes have been archived and maintained as a centralized institutional resource. MBARI has developed a set of software applications to annotate and access video data. Users can identify the location of video sequences using a data query component; complex queries can be made by constraining temporal, spatial, or physical parameters (e.g., season, location, or depth). The applications reference a knowledge base of over 3,000 biological, geological and technical terms, providing consistent hierarchical information about objects and associated descriptions for annotating video at sea or on shore. The annotation, knowledge base, and query components together provide a comprehensive video archive software system that can be applied to a variety of scientific disciplines. Also in development, using the XML data format, is an interactive reference interface to explore MBARI's deep-sea knowledge base. When complete, the full software system will be disseminated to the research community via the web or CD, to help meet the challenges inherent in archiving video data. http://www.mbari.org

[back to program schedule]

OS61C-09; INVITED; EOS, Trans. AGU, 83 (47), F700

Interoperability Among Spatial Data Resources Along a Continuum: "Data-to-Data", "Data Models" and "Data-to-Interpretation"

Wright, D J
dawn@dusk.geo.orst.edu
Oregon State University, Department of Geosciences 104 Wilkinson Hall, Corvallis, OR 97331-5506 United States

While there is clearly a wealth of digital and analog data from the oceans to facilitate interdisciplinary studies, the diversity, volume, and complexity of these data sets makes it extremely difficult to efficiently and intelligently transform the data into information, and further to synthesize knowledge from studies over large geographic areas. Presented is a brief overview of four projects in progress that seek to promote the interoperability of data and software to, in turn, facilitate the leap from scientific data access to knowledge discovery. The projects range in their focus from a system with simple access to metadata and data, as well as linkages between disparate data sets (data to data), to a standard object-oriented data model for the structure of databases with "rules" for behavior and placeholders for analytical functions, to a complex computational environment (consisting of web mapping, relational database management, and analytic tool composition by the user) that facilitates refinement of numerical simulations, quantitative evaluation of scientific hypotheses, and exploration of new relationships between observables (data to interpretation).

[back to program schedule]

OS61C-03; EOS, Trans. AGU, 83 (47), F698

User-Friendly Data Servers for Climate Studies at the Asia-Pacific Data-Research Center (APDRC)

Yuan, G; Shen, Y; Zhang, Y; Merrill, R; Waseda, T; Mitsudera, H; Hacker, P
yuan@soest.hawaii.edu
IPRC-SOEST, Univ. of Hawaii, 1680 East West Rd. POST 401, Honolulu, HI 96822

The APDRC was recently established within the International Pacific Research Center (IPRC) at the University of Hawaii. The APDRC mission is to increase understanding of climate variability in the Asia-Pacific region by developing the computational, data-management, and networking infrastructure necessary to make data resources readily accessible and usable by researchers, and by undertaking data-intensive research activities that will both advance knowledge and lead to improvements in data preparation and data products. A focus of recent activity is the implementation of user-friendly data servers. The APDRC is currently running a Live Access Server (LAS) developed at NOAA/PMEL to provide access to and visualization of gridded climate products via the web. The LAS also allows users to download the selected data subsets in various formats (such as binary, netCDF and ASCII). Most of the datasets served by the LAS are also served through our OPeNDAP server (formerly DODS), which allows users to directly access the data using their desktop client tools (e.g. GrADS, Matlab and Ferret). In addition, the APDRC is running an OPeNDAP Catalog/Aggregation Server (CAS) developed by Unidata at UCAR to serve climate data and products such as model output and satellite-derived products. These products are often large (>2 GB) and are therefore stored as multiple files (stored separately in time or in parameters). The CAS remedies the inconvenience of multiple files and allows access to the whole dataset (or any subset that cuts across the multiple files) via a single request command from any DODS enabled client software. Once the aggregation of files is configured at the server (CAS), the process of aggregation is transparent to the user. The user only needs to know a single URL for the entire dataset, which is, in fact, stored as multiple files. CAS even allows aggregation of files on different systems and at different locations. Currently, the APDRC is serving NCEP, ECMWF, SODA, WOCE-Satellite, TMI, GPI and GSSTF products through the CAS. The APDRC is also running an EPIC server developed by PMEL/NOAA. EPIC is a web-based, data search and display system suited for in situ (station versus gridded) data. The process of locating and selecting individual station data from large collections (millions of profiles or time series, etc.) of in situ data is a major challenge. Serving in situ data on the Internet faces two problems: the irregularity of data formats; and the large quantity of data files. To solve the first problem, we have converted the in situ data into netCDF data format. The second problem was solved by using the EPIC server, which allows users to easily subset the files using a friendly graphical interface. Furthermore, we enhanced the capability of EPIC and configured OPeNDAP into EPIC to serve the numerous in situ data files and to export them to users through two different options: 1) an OPeNDAP pointer file of user-selected data files; and 2) a data package that includes meta-information (e.g., location, time, cruise no, etc.), a local pointer file, and the data files that the user selected. Option 1) is for those who do not want to download the selected data but want to use their own application software (such as GrADS, Matlab and Ferret) for access and analysis; option 2) is for users who want to store the data on their own system (e.g. laptops before going for a cruise) for subsequent analysis. Currently, WOCE CTD and bottle data, the WOCE current meter data, and some Argo float data are being served on the EPIC server. http://apdrc.soest.hawaii.edu/

[back to program schedule]