ArcOD Data

What is the Arctic Ocean Diversity Project (ArcOD)
The Arctic Ocean Diversity Project of the Census of Marine Life (http://www.coml.org/descrip/aobio.htm) is an international collaborative effort to inventory biodiversity in the Arctic’s sea ice, water column and sea floor from the shallow shelves to the deep basins. It employs a three-layer approach: compilation of existing data, taxonomic identification of existing samples, and new collections focusing on taxonomic and regional gaps. ArcOD contributes to OBIS (Ocean Biogeographic Information System; www.iobis.org/), and is housed within SFOS (School of Fisheries and Ocean Sciences)/IMS (Institute of Marine Sciences) at the University of Alaska-Fairbanks, USA. Russian partner centers are located at the Zoological Institute in St. Petersburg and at the Shirshov Institute of Oceanology in Moscow, Russia. ArcOD was formed in 2004, and part of the CoML/OBIS scheme until at least 2010.

What Data are in ArcOD
ArcOD serves the global village with marine species data from the circumpolar Arctic (http://ak.aaos.org/op/data.php?region=ARC&name=obis). Datasets in ArcOD encompass sea ice algae and phytoplankton as well as zooplankton, zoobenthos, fish and birds. ArcOD currently serves more than a dozen datasets, Metadata included, and has over 50 datasets identified for a work up, and coming forward over the year. These precious data are contributed from >7 nations, and were derived from museum collections, historic and recent Arctic research expeditions.

What are the technicalities within ArcOD
ArcOD is using a distributed online network structure following the popular DiGIR protocol (z39.50 compliant; http://www.cni.org/pub/NISO/docs/Z39.50-brochure/). ArcOD runs LINUX servers and is compatible and registered with OBIS and GBIF (Global Biodiversity Information Facility www.gbif.org). The ArcOD database is housed with the Alaska Ocean Observing System (AOOS; www.aaos.org), and is therefore searchable in all relevant global and national search engines, e.g. USFirstGov (http://www.usa.gov/), and NBII Clearinghouse (http://mercury.ornl.gov/nbii/). This allows for global access of ArcOD data, as well as for exposure of data contributors. A sophisticated data back-up system is also in place.

Who recognizes ArcOD
Beyond the mentioned institutions such as OBIS, GBIF and AOOS, ArcOD is also the cluster lead of the Arctic marine biodiversity cluster of the IPY (International Polar Year) 2007/8 (http://classic.ipy.org/development/eoi/proposal-details.php?id=333).

Why do FGDC NBII Metadata matter for ArcOD
Metadata are a key component in the ArcOD concept, because one cannot use data unless they are well documented. ArcOD follows one of the most sophisticated and detailed Metadata Standards, FGDC (Federal Geographic Data Committee; http://www.fgdc.gov/metadata) NBII (http://www.nbii.gov/portal/server.pt), thereby assuring global information availability with the highest possible quality.

Where ArcOD is going
By 2010, ArcOD will provide more information and data to the public. Besides data, it also provides species pages including images, maps, outreach materials, and synthesis products. In order to provide even more for a truly global approach, ArcOD is planning to link with the SCAR-MarBIN Antarctica databases, and implement their Webportal for consistency. It is our explicit goal that ArcOD will become a globally recognized information provider to better facilitate addressing climate change-related questions and sustainable decision-making for the Arctic and beyond.

For more information, please contact: the ArcOD team (arcod@ims.uaf.edu)
British Oceanographic Data Centre (BODC)

Data type(s) that are managed within your system

BODC is the UK National Oceanographic Data Centre in the IODE network and the Natural Environment Research Council’s designated data centre for marine data. The NERC remit requires handling all types of marine data whose collection has been funded by NERC that are not explicitly covered by other NERC centres. This exception currently includes remotely sensed data, model output and marine geology/geophysics data but BODC’s remit, particularly in the areas of geological data and model output is likely to expand.

The majority of BODC’s data holdings are observational data in the saltwater column including thousands of physical, chemical, biological and geological parameters and associated measurements such as meteorology. We also hold smaller amounts of soft core, freshwater and atmospheric data acquired as part of multidisciplinary project data sets.

BODC is primarily a ‘delayed mode’ data centre with a strong emphasis on data and metadata quality assurance during ingestion. However, management of operational data acquired in near real time (UK Argo floats, Rapid Climate Change moorings and CarbonOPS ship systems) are assuming increasing importance.

BODC is the centre responsible internationally for the management of digital bathymetric data on behalf of GEBCO as described on the GEBCO page.

Scope/mandate of the system (e.g. programmatic, spatial, or thematic scope)

The BODC data system may be regarded as multi-faceted with data accessible through thematic metadata, programmatic metadata, activity (such as cruises) metadata or through exposure of their spatio-temporal co-ordinates.

Data policies (e.g. access restrictions)

BODC assembles datasets collected in the field by others and therefore operates a wide range of data policies negotiated with the data originators. NERC data are covered by an organisational data policy. Data policies may be highly restrictive for data supplied by the military or commercial sectors, but the objective of the data centre is to bring data into open use for the education and research sectors as quickly as possible. The commercial sector is charged on a ‘handling cost recovery’ basis for data access.

Access interfaces & tools (including web services, APIs, etc.)

There are a range of data delivery systems available through the BODC Web Site (http://www.bodc.ac.uk/data/online_delivery/) and further systems are under development. Some of these deliver data directly, whilst others build a ‘shopping basket’ for subsequent manual request handling. The strategic objective is to increase the proportion of data delivered automatically whilst honouring the complex network of data policy agreements.

BODC is involved in two distributed data system projects (NERC DataGrid and SeaDataNet) that reach the pre-operational stage in the next 12 months.

Metadata standards, vocabularies

Both NERC DataGrid and SeaDataNet have adopted a standards-based approach founded on the outputs of OGC and ISO TC211 (such as ISO19115/19139) and these are being adopted wherever possible by BODC. Content and technical governance has been established jointly by SeaDataNet and the IOC MarineXML Steering Group to develop the vocabularies required for profiling of the core standard. Vocabularies are served through a Web Service API and client interface (http://vocab.ndg.nerc.ac.uk/client/vocabServer.jsp).
ChEssBase – a web-based relational database for species from deep-water chemosynthetic ecosystems integrated with OBIS.
Eva Ramirez-Llodra & Maria Baker
National Oceanography Centre, Southampton, UK

ChEssBase is a dynamic relational database available online since December 2004. The aim of ChEssBase is to provide taxonomical, biological, ecological and distributional data of all species described from deep-water chemosynthetic ecosystems, as well as a bibliography and information regarding the habitats. These habitats include hydrothermal vents, cold seeps, whale falls, sunken wood and areas of minimum oxygen that intersect with the continental margin or seamounts. ChEssBase is a data source of OBIS, the Ocean Biogeographic Information System. OBIS, the information component of the Census of Marine Life (www.coml.org) is a web-based dynamic provider of global geo-referenced information on marine species in four dimensions (the three dimensions of space plus time). OBIS Data Sources are databases made available through OBIS.

1- Data type(s) that are managed within your system
Biological data including full taxonomy of species, basic biological and ecological data, distribution (LAT, LON, depth), endemicity and relevant publications. Habitat information including site location, basic geological characteristics and relevant publications. In 2007, ChEssBase will include data on specific biological samples, where, when and how they were collected, who holds them.

2- Scope/mandate of the system (e.g. programmatic, spatial, or thematic scope)
ChEssBase has been designed and built in the framework of the ChEss-Census of Marine Life project for all species from deep-water chemosynthetic ecosystems, including hydrothermal vents, cold seeps, whale falls, sunken wood and OMZs. However, the database is open to all researchers/institutes wishing to use it as a depository of geo-referenced biological data on species from deep-water chemosynthetic ecosystems.

3- Data policies (e.g. access restrictions)
ChEssBase is open to all users through a search facility online. There are no access restrictions. Users are only asked not to use ChEssBase for commercial purposes and to acknowledge the original data sources (original publications) and ChEssBase in analyses conducted using the ChEssBase facility.

4- Access interfaces & tools (including web services, APIs, etc.)
ChEssBase is available online, with a description page and a search interface: http://www.noc.soton.ac.uk/chess/database/database.html
ChEssBase has been integrated with OBIS (Ocean Biogeography Information System) since 2005 and therefore the data are also available through the OBIS search interface: http://www.iobis.org/

5- Metadata standards, vocabularies
Data Workshop participant information

Participants from the Continental Shelf Programme, UNEP/GRID-Arendal, Norway:
Yannick Beaudoine, PhD, Project Geoscientist
Øystein Halvorsen, MSc, Senior Systems Consultant
Tina Schoolmeester, MSc, Project Geoscientist

1. Scope/mandate of the system:
The UNEP Shelf Programme at UNEP/GRID-Arendal in Norway has developed a One Stop Data Shop (OSDS) for use by coastal states preparing submissions for an extended continental shelf beyond 200 nautical miles. The OSDS was developed in response to a United Nations resolution. The OSDS aims to efficiently provide the most complete marine geoscientific data package for a State’s Article 76 area of interest. Additional OSDS capacities include data manipulation, modelling, visualization and storage.

2. Data type(s) that are managed within our system:
- Cruise metadata
- Cruise track lines w. corrected bathymetry, grav/mag observations, date/time information
- Global bathymetry and sediment thickness grids (modelled data for preliminary assessment purposes)
- Multibeam data
- Digital seismic profiles (e.g. SegY format)
- In house processed and georeferenced analog seismic profiles
- Integrating sediment velocity data to provide accurate sediment thicknesses
- Geological sample information and core log data (e.g. ODP, DSDP and IODP)

3. Data policies:
Distribution policies adhere to the requirements and restrictions outlined by the different data holders and requesting states. No data will be made available online and only official request from qualifying states (developing states and small island states) will be considered.

4. Access interfaces & tools
States preparing an Article 76 submission can access the OSDS’ data inventory map online at http://maps.continentalshelf.org/ The online tool uses an ArcIMS application showing stick maps and metadata for the cruise track lines that we have in house with references to the respective data owners.

Our main application tool is Geocap (http://www.geocap.no) equipped with a particular UNCLOS module. All data are stored in VTK format with numerous exporting format capabilities. A current OSDS project is the development of an SQL-based solution for appropriate handling of metadata in Geocap.

Data conversion is done by standard tools like GMT, MBsystem, Freeform and ArcGIS. In addition a suite of programs has been internally developed using CSV files as the generic format for vector and point data. The suite can translate to/from VTK, ESRI Shapefile, MBsystem mblst, DXFMT, MGD77, Dif, miscellaneous XYZ(T) and nav/dat files, as well as combine or split CSV files, manipulate fields and names, check formats, weed and insert line breaks to create stick maps, convert zone (polygon) files into border (line) files, etc.

5. Metadata standards, vocabularies
This document has been prepared for the MARGINS workshop in Kiel, May 2007, in response to the call for information on data repositories.

Preamble

The EarthLook system is in its early stage. It will be accessible via www.earthlook.org.

1. Data types managed

Following our research main stream, the central data type is raster data. We are addressing multi-dimensional rasters, such as 2-D seafloor maps and satellite maps, 3-D x/y/t satellite image time series and x/y/z seismic data, and 4-D x/y/z/t climate models.

Additionally catalog and vector data will be added, depending on availability and relevance to the users.

2. Scope/mandate

EarthLook represents a breadboard for a GMES Center Bremen, focusing on a maritime theme. This includes all kind of raster data appearing, such as (but not limited to): satellite imagery (hyperspectral, radar, …); oceanographic and atmospheric simulation models; seafloor observation data (such as ROV video); bathymetry data; seafloor seismic data; etc.

3. Data policies

Unfortunately it is not easy to obtain an unconditional clearance from the various (institutional and scientific) data providers, for various reasons. We keep on working on this and hope that a public access policy can be achieved soon.

4. Access interfaces & tools

EarthLook is devoted to open, standards-based interfaces. The EarthLook portal will offer access via OGC WMS, WCS, WCPS, eventually WFS and OGC catalog services. WCPS, in particular, will allow the user to phrase requests of any complexity against raster data archives.

5. Metadata standards, vocabularies

none so far, as our main interest is on value-added raster services.
Earthquake Research Center (EQRC), Ferdowsi University and Khorasan Seismic Network

Hossein Sadeghi, S.Keivan Hosseini, M.Mahdi Motiei  
Ferdowsi Univ., Mashhad 91779-48974, Iran  
Tel: (+98-511) 8791195, Fax: (+98-511) 8791191

- **Data types:**
  - Full Frequency (125 sps) .eve and ASCII format
  - Short Period (25 sps) .eve and ASCII format
  - Long Period (1 sps) .eve and ASCII format

- **Scope/mandate of the system:**
  - Khorasan Seismic Network was designated to record the earthquakes in all over Khorasan provinces (North, Razavi and South Khorasan) and neighbour area - 500 km and 1000 km in radius for local and regional earthquakes, respectively. The network also records the earthquakes larger than M6 in all around the world. There are eight stations in the network:
    1. Bojnurd
    2. Quchan
    3. Mashhad
    4. Sabzevar
    5. Najafi
    6. Ferdows
    7. Kakhk
    8. Birjand

- **Data policies:**
  - Our data are ready to use for research and academic activities. Requests should be received via e-mail (sadeghi@seismo.um.ac.ir) at the present time. It is also possible to access data via our web site (http://seismo.um.ac.ir/) using username and password for members.

- **Access interfaces and tools:**
  - At the present time access to the seismic stations and data transferring are possible via dial-up modems.
  - Seismometers:
    - Guralp CMG-3T 120s-50Hz frequency response
  - Digitizers:
    - RAU (ISMES Co. digitizer)
    - CMG-DM24
General Bathymetric Chart of the Oceans (GEBCO)

GEBCO makes available global ocean floor bathymetric data sets. It operates under the joint auspices of the International Hydrographic Organization (IHO) and the Intergovernmental Oceanographic Commission (IOC). http://www.gebco.net

Data type(s) that are managed within your system

GEBCO’s data sets consist of a one arc-minute bathymetric grid (the GEBCO One Minute Grid), bathymetric contours and accompanying trackline control; printed bathymetric charts and a gazetteer of geographic names of undersea features. The bathymetric contours and GEBCO One Minute Grid are included as part of the GEBCO Digital Atlas (GDA). The GDA is maintained and updated on behalf of GEBCO at the British Oceanographic Data Centre.

Scope/mandate of the system (e.g. programmatic, spatial, or thematic scope)

The data are spatial, covering one theme, i.e. bathymetry.

Data policies (e.g. access restrictions)

The GEBCO One Minute Grid is available free of charge via the web. There is a charge made for the GEBCO Digital Atlas CDROM. The data sets within the GDA are distributed under the condition that data are not passed on to third parties without prior permission and that attribution is given to the source of the data sets.

Access interfaces & tools (including web services, APIs, etc.)

The GDA is available on a CDROM, which includes a software interface for viewing and accessing the data. The GEBCO One Minute Grid is also available online via a web application (http://www.bodc.ac.uk/data/online_delivery/gebco/). Software can be downloaded from the web for viewing and accessing the gridded data. The gazetteer of geographic names of undersea features is also available for download via the web.

P.W. 27/04/2007
GeoForschungsZentrum Potsdam

Jens Klump, Doris Dransch, Anastasia Galkin

GeoForschungsZentrum Potsdam (GFZ) is a non-university geoscientific research institute. It was founded on January 1st, 1992 on the Telegrafenberg in Potsdam. As the first of its kind worldwide, GFZ Potsdam combines all solid earth science fields including geodesy, geology, geophysics, mineralogy and geochemistry, in a multidisciplinary research centre.

Research is accomplished by the use of a broad spectrum of methods and techniques, such as satellite geodesy and remote sensing, geophysical deep sounding, scientific drilling, experiments under in-situ conditions and modelling of geo-processes.

The GFZ maintains various instrument pools for field research and global measurement campaigns, a team of engineers for the development of geoscientific instruments and a group of specialists for the Task Force Earthquake. An underlying principle is to combine the geoscientific know-how of universities and other research centres in national and international joint projects.

Data Management

GFZ Potsdam provides data management for several international projects in earth science. Among these are the International Continental Scientific Drilling Program (ICDP), on-site data management for Mission Specific Platforms of the Integrated Ocean Drilling Program, storage and distribution of data products from the CHAMP and GRACE satellite missions, the GEOFON seismological network, and others. The Data Center at GFZ Potsdam is a publication agent for data publications through the STD-DOI system.

Sensor Networks

In many projects GFZ Potsdam operates sensor networks and geophysical observatories. Among these are GPS arrays, seismological stations, geomagnetic observatories, tide gauges, and others. The most extensive of these networks serves the German Indonesian Tsunami Early Warning System. Current research at the GFZ Potsdam Data Center investigates technologies to integrate sensor networks, data processing and data distribution through the use of standardised interfaces between these components.

Geovisualization and Knowledge Management

This group is active in the following fields:

• Development of applications to access information and scientific knowledge in models and methods via the internet. The applications are based on Service oriented architecture (SOA) and workflow engines. As an example, we have implemented a SOA application to access a dike breach model. This application combines several
standardized OGC- and web-services which are orchestrated by a BPEL workflow engine.

- Presentation of context- and user-oriented information for different views on information. We work with task analysis, role concepts and salient visualization to realize this.
- Development of a cross-community collaborative visual platform for comprehensive understanding of the subduction mega-earthquake mechanisms and the related seismological structure with focus on the Andes region. This platform will combine methods from knowledge and information representation with geo-visualizatio methods.
GEOROC and GeoReM Databases – Linking Chemical Data and Data Quality

B. SARBAS, K.P. JOCHUM, U. NOHL
Max-Planck-Institut für Chemie, Postfach 3060, 55020 Mainz, Germany; sarbas@mpch-mainz.mpg.de

Since its introduction about eight years ago, the geochemical database GEOROC (Geochemistry of Rocks of the Oceans and Continents, http://georoc.mpch-mainz.gwdg.de) of the Max-Planck-Institut für Chemie in Mainz established itself as a major online resource available to the scientific community. GEOROC provides geochemical data published for volcanic whole rocks, glasses, minerals and inclusions from ocean islands, large igneous provinces, convergent margins, Archean greenstone belts, rift and intraplate volcanic regions. The database now provides about 275,000 analyses published in about 6,500 papers (status: 02/2007).

The web interface of GEOROC allows the selection of samples by bibliographic, tectonic, geographic, petrological as well as chemical criteria. As part of the bibliographic query, the search for the GEOROC_Reference_Number permits an easier reproduction of published data plots created with the help of the database. A new query based on age information will be added in the near future.

The geochemical database GeoReM (Geological and Environmental Reference Materials, http://georem.mpch-mainz.gwdg.de) for reference materials and isotopic standards includes reference samples from rock powders, glasses, minerals, isotopic standards, river water, and seawater. GeoReM is a relational database with a structure compatible to that of GEOROC and contains published analytical and compilation data (major and trace element concentrations, radiogenic and stable isotope ratios), important metadata about the analytical values, such as uncertainty, uncertainty type, method and laboratory. Sample information and references are also included.

Five different queries are now possible: Samples or materials, GeoReM preferred values, chemical criteria, bibliography, and methods and institutions. GeoReM contains more than 1200 geological reference materials and 10000 analyses (status: 12/2006).

References available in both databases are linked to each other. It is now possible to get information through GEOROC about the geochemical analyses of rock samples belonging to analyses of reference materials and isotopic standards compiled in GeoReM. The detailed information of the analytical conditions available in GeoReM enables users of GEOROC to estimate the quality of the analysed rock samples.

In spring 2003, GEOROC joined with the databases PetDB and NAVDAT to initiate the EarthChem consortium (http://www.earthchem.org/), with the aim to increase the synergy between the three geochemical database efforts. The recently introduced EarthChem Portal (http://geoportal.kgs.ku.edu/earthchem/jtest/) offers a seamless search across the three databases using reference, location or rock type as selection criterion.
GNS Science - Data Facility Description

Craig Jones
Data Administrator - Hydrocarbons

GNS Science is a Crown Research Institute (CRI) wholly owned by the New Zealand Government and incorporated on 1 July 1992.

Its role is to be the government's principal earth systems and isotope science researcher, providing expertise on off-shore and on-shore energy and mineral resources, geological hazards and risk, and their economic, social, and environmental impacts.

As such it holds a wide variety of geological, geochemical & geophysical information from around the New Zealand region including passive & active seismic, volcanological, geothermal, groundwater, hazards and engineering datasets.

Underpinning the storage and management of these datasets are a number of core GNS Information Systems. They are:

PETLAB - The New Zealand, Southwest Pacific and Antarctic rock catalogue and geoanalytical database, operated by GNS with contributions from most of New Zealand's other earth science research institutions (primarily the Universities). PET (GNS and university rock catalogues) data can be queried and downloaded by anyone free of charge. Access to LAB (geochemical, isotopic, age, volumetric and petrophysical) data requires free registration.

GERM - the GEological Resource Map of New Zealand database is a mineral deposit inventory of mineral, rock aggregate, building stone, coal, oil, gas, thermal water, and cold-water spring occurrences. More than 10,000 sites such as quarries, mines, dredges, wells, outcrops, seeps, springs, or fields are catalogued.

GGW - contains geochemical (chemistry and sampling), hydrological, geological and/or geophysical information for sites in New Zealand. Included in the database are the datasets for the National Groundwater Monitoring Programme. Projects are maintained and operated by GNS Science.

FRED - New Zealand's Fossil REcord File, a database for fossil localities in New Zealand and nearby regions including Southwest Pacific Islands and the Ross Sea region of Antarctica. It is jointly managed by Geological Society of New Zealand and GNS Science on behalf of, and for the benefit of, all New Zealand Earth scientists. Full use of the database requires free registration.

GEONET – This project is a collaboration between the New Zealand Earthquake Commission and GNS Science to provide real-time monitoring and data collection for rapid response and research into earthquake, volcano, landslide and tsunami hazards. All data and images are made available free of charge to facilitate research into hazards and assessment of risk.

PSID – The Petroleum Systems Database which provides an integrated catalogue for the management of GNS's holdings of available petroleum exploration (well and seismic) datasets for the whole of New Zealand. Limited external access to data is available via a spatial catalogue of publicly available data.

QMAP – This GIS database is the underpinning data management system for the program to remap New Zealand at 1:250 000 scale. It will ultimately be integrated into a seamless national geological database accessible to the New Zealand public in vector and raster formats.

All of these core systems are housed in a GNS corporate Oracle database and access is primarily via the web using java applets developed in-house and ESRI SDE via ArcIMS web services.

Supplementing these systems are extensive geophysical datasets (marine geophysical cruises, research and petroleum exploration seismic) that are currently stored in proprietary software Schlumberger's Geoquest) and on digital media (tapes, DVDs, CDs) and catalogued in a digital media catalogue. Future plans include expansion of the existing information systems to include these datasets.
Geophysical observation network of IFREE/JAMSTEC in Northwest Pacific Region
Seiji Tsuboi, Yasushi Ishihara, Hiromitsu Mizutani, Hiromichi Nagao, Yasuko Yamagishi and Yoshiki Yuki
IFREE/JAMSTEC

Institute for Research on Earth Evolution (IFREE) of JAMSTEC has maintained the geophysical observation network in western Pacific region, which were installed by the Ocean Hemisphere Network Project during late 90s. There are now 10 broadband seismograph stations and 7 geomagnetic stations in this network. Most of the broadband seismograph stations are equipped with three component STS-1 broadband seismometers. Geomagnetic stations are equipped with both fluxgate and proton magnetometers and observed data are sent to the IFREE data center by postal mail. The data recording system at each broadband seismograph station has been upgraded to Quanterra Q330 data logger after the stations are transferred to IFREE. We use Antelope software to retrieve broadband seismograms recorded at each station in realtime to IFREE data center in collaboration with the Earthquake Research Institute of the University of Tokyo. Figure 1 illustrates distribution of broadband seismograph station. The data obtained through this network are archived in the data center (http://www.jamstec.go.jp/pacific21/) and opened to researchers after quality control of data. These data are also available through Geophysical Data Service client software, which are developed using Web Services technologies. The software is also available through the web site of our data center.

Fig. 1 Broadband seismograph stations operated by IFREE/JAMSTEC.
In the research of Earth sciences, it is recognized that we need to compare various types of data, such as geophysical, geological, and geochemical data to have new images of the Earth’s current status. There has been no way to compare these data directly using some database viewing tools. In 2005, new software called Google Earth has started its service. Google Earth enables us to view satellite image data for any regions of the Earth's surface. It also has a feature to visualize local data, which exists on user's PC, on top of the satellite images. This can be realized by converting user's data to KML (Keyhole Markup Language), which is an XML-based language. Many geophysical and geological data are expressed by its location (latitude and longitude) and its value, such as gravity anomaly, geomagnetic anomaly, elevation, rock types etc. Thus, the Google Earth should become a useful tool to handle geophysical and geological data and compare various kinds of data. We have developed software to convert these digital data into KML format and verified that the data can be viewed successfully by the Google Earth (Fig. 2). We are now developing software to retrieve geochemical data through existing database, such as PetDB, and convert them to KML format. This software will allow us to overlay geophysical, geological and geochemical data for some specified region of the Earth, which has never been possible before, and will enable us to have new images toward the interaction of various disciplines of Earth system.

Fig. 2 Seismic velocity distribution under Japanese Islands overlaid by the satellite imagery using the Google Earth.
Japan Agency for Marine-Earth Science and Technology (JAMSTEC) operates five research vessels and three manned/unmanned submersibles for oceanographic research activity. Data and samples obtained on its research cruises/dives are basically opened to public through JAMSTEC website (http://www.jamstec.go.jp/e/index.html). You will find about a dozen of data sites in its Database page. Some data sites are operated by MEDID and others are by respective research centers.

1. Data Types
JAMSTEC research activity extends in physical/chemical oceanography, biology, meteorology, geology, seismology etc. and from observation to prediction. JAMSTEC data sites also provide photographs and video images taken by submersibles and metadata of samples also taken by them. Data format depends on the data sites.

2. Scope
JAMSTEC data are mainly obtained in the area of northwestern Pacific and also include other Pacific regions, Indian Ocean, Arctic and Antarctic Oceans, Atlantic. Data are from 1980s starting operation of research vessels to now, but old data is very limited.

3. Data Policies
Comprehensive data policy is recently established (effective for FY2008 cruises) instead for provisional data policies. In this data policy (i) JAMSTEC basically owns all data/samples obtained through its activity, (ii) Researchers obtained data/samples can use them exclusively for two years, (iii) JAMSTEC make data/samples open to public after that period free for scientific use. MEDID is to be the data center in JAMSTEC. As of now it could have opened partially the data stored in JAMSTEC.

4. Access Interface & Tools
Data sites are running on web servers, so users get information via web browsers. Some sites may be designed for specific versions of browsers and request for java script.

5. Metadata Standards, Vocabularies
JAMSTEC does not have its metadata standard yet and each data site uses its own metadata. MEDID is planning to establish JAMSTEC original metadata standard based on existing metadata standards.
Ifremer develops and operates major facilities to collect, manages and archives data for a wide variety of topics. Users are now requiring interfaces which provide transparent access to multidisciplinary datasets which are distributed among several existing systems operated in different locations by several teams. To meet their needs, Ifremer decided as NODC to build a data portal, Nautilus, supplying a unique entry point to all the users, with a main goal: to build it on existing systems without changing them. Nautilus is opened since June 2006:

http://www.ifremer.fr/nautilus

**Data types**

Nautilus aims to give an access to all kinds of data: biodiversity, ecosystems, ocean circulation; exploration of the ocean floor... At present the access is given to in situ data of chemistry, of physical oceanography and of geophysics. It is being widened to new types of data such as geology. The system will be spread to observation data relating to environment, to fisheries and to satellite data (multi-mission synthesis).

**Scopes**

The scopes are:
- to give an integrated access to all kinds of data to support multidisciplinary programs: fundamental research as well as operational oceanography in response to the request of the public actors and according to the international commitments of the State (environment and climat monitoring and forecasting, integrated management of coastal zone)
- to ensure the interoperability of the Ifremer systems with the national and european ones (Seadatanet)

**Data policies**

Data are available according to the IOC recommandations, the UNCLOS, the European directives, the French Law, the Ifremer (chief scientist agreement required for recent surveys) and partner regulations. Within EEZ, restrictions may apply – especially to geosciences data. For industrial and commercial purposes, data are made available at cost. Any use of the data should acknowledge their source according to the Ifremer copyright

**Access interface and tools**

Nautilus provides a web based user interface with the following user functionalities:
- Data discovery: Where? What? How (observation conditions)? When? Who
- Visualization of observation locations (mapping tool)
- Online data access

**Metadata standards, vocabularies**

Considering that all existing data management systems share common requirements based on geographical positioning of observations, standards related to geographic information are extensively used as the basement of this portal: ISO 19115 standard to describe data sources and datasets and OpenGIS standards to visualize dataset geographical locations. In addition, standards like NetCDF data format, OpenDAP data access protocol, SOAP for web-based services are used to network the existing systems and transport users queries and results. Used vocabularies are developed and maintained at the European level within the Seadatanet program.
1. Data type(s) that are managed within your system
The IRIS Data Management Center (DMC) primarily manages time series data. The majority of these data are seismological time series but approximately 24 different kinds of time series data are available from globally distributed sensors including weather data, hydrologic data and other types of geophysical data such as gravimeter data and magnetotelluric data.

More recently the DMC has developed systems to manage products derived from observational data, in essence any arbitrary collection of information can now be managed, queried for and distributed to an end user,

2. Scope/mandate of the system (e.g. programmatic, spatial, or thematic scope)
The IRIS DMC mandate is programmatic. We are charged by NSF to manage all data generated by the IRIS Global Seismic Network, the Program for Array Seismic Studies of the Continental Lithosphere (PASSCAL) and other seismological data to guarantee easy access to other sources of seismic data including FDSN data and data generated by the USGS. We are a data center that manages global (and beyond) time series data.

3. Data policies (e.g. access restrictions)
Data generated by individual scientists as a result of PASSCAL type deployments are restricted to the PI and those the PI authorizes for a period of two years after the experiment is demobilized. If these experiments have more that 5 instruments deployed at least one of these stations must be open and data made available without any artificial delay. Some other international projects such as the OBS Instrumentation Program and SEIS-UK have separate data restriction policies that IRIS enforces.

All other data managed by the IRIS DMC is available, often in real time, without any artificial delay. Data are available without charge and available to everyone.

4. Access interfaces & tools (including web services, APIs, etc.)
The IRIS DMC has a variety of tools available through which users can make requests for data from its holdings. These include email based, browser based, CORBA based API's and the DMC is now developing SOAP and REST based web services as well as web mapping services.
An introduction to data request tools at the IRIS DMC are available at http://www.iris.edu/data/data.htm

A tutorial on how to gain access to the data holdings at IRIS can be found at http://www.iris.edu/manuals/DATutorial.htm

A simple access figure allowing quick access to data and request tools is found at http://www.iris.edu/data/req_methods.htm

A request tool wizard to help you decide which request tool is best suited to your data needs can be found at http://www.iris.edu/data/wizard/tools.htm

A data wizard, in development, that can help you determine what kind of data the IRIS DMC has available can be found at http://www.iris.edu/data/wizard/data.htm

5. Metadata standards, vocabularies

The metadata standard that IRIS and the FDSN uses to describe seismic stations and seismograms comes from the Standard Exchange for Earthquake Data (SEED) standard developed by the FDSN an IASPEI/IUGG/ICSU commission.

The IRIS DMC uses standards such as the OGC Web Mapping Services where applicable, as well us Dublin Core and a variety of XML including QuakeML.

There is no significant vocabulary effort being pursued in the FDSN or IRIS at this time.
Marine Seismic Data Center at the University of Texas Institute for Geophysics

Thomas H Shipley, Lisa Gahagan, Kevin Johnson and Marcy Davis

The Marine Seismic Data Center (MSDC) provides free access to seismic images and data for education and research. It is designed to improve sharing of digital seismic data from U.S. National Science Foundation-sponsored research and other sources. In collaboration with Lamont-Doherty Earth Observatory (LDEO), the University of Texas Institute for Geophysics (UTIG) developed a structured method to preserve digital data, simplifying discovery, access and delivery to the academic-scientific community. This has resulted in a pair of data centers, the MSDC described here and the LDEO Field Data Center (FDC), which are linked by shared metadata and interoperability. Both are part of the Marine Geoscience Data System, providing integrated access to public data (www.marine-geo.org).

Scope. The MSDC contains processed sections from over 170 cruises, growing as legacy data is incorporated from individual scientists, other programs and from the LDEO archive. The system totals about 25,000 primary data files of which 8,000 are stacks, migrations, single-channel and ocean bottom seismometer lines. The remainder are field data from 85 cruises, primarily from UTIG ship operations prior to 1988. Images of seismic profiles and associated navigation are produced for all data, providing a reach resource of education research. Extensive metadata provide building blocks for searching the data inventory.

Access interfaces. We rely on a server-side web interface for most access. Metadata searches produce tables with pre-selected parameters. Map-based searches locate seismic images for inspection, with figure captions, links to project or cruise organized information, contributors, references, navigation, and seismic data. Track lines may be filtered for a particular data type (stacks, migration, shots, OBS, etc). An online seismic plotter, using SIOSEIS (band-pass filter, gain and plot), produces user-customized raster and postscript files. A download cart collects selected files, which are extracted from the disk/tape library for downloading through a web page link.

Web services provide another mechanism for sharing metadata. Location information is hosted as an Open Geospatial Consortium-compliant web mapping service. However, the WMS mapping service is not generally useful because of the large navigation volumes (3 GB). We also produce a static file in Google Earth (kmz) that provides a far better user experience.

Metadata standards. The metadata follow the Federal Geographic Data Committee standards and developing ontologies for marine and seismic data. As part of the Marine Geoscience Data System, we are following ongoing efforts including the Marine Metadata Interoperability project. Developing true interoperability with the LDEO Field Data Center (with REST) tests the metadata structure, convergence of relevant metadata and methods of linkage and communication.
What is the National Geophysical Data Center (NGDC)? NGDC is the steward of a wide variety of global geophysical data, maintaining the world's largest collection of seafloor and space environment data, and other geophysical observations such as geomagnetism. It is also the sole archive of data from the US Department of Defense’s Defense Meteorological Satellite Program (DMSP).

NGDC provides critical services. Scientists and policy makers require access to high quality geophysical and space environment data to improve our understanding of the Earth and near-Earth environment, the impact of solar-terrestrial processes on population and property, and the state of the environmental, and economic security. NGDC is the data archive of NOAA’s Space Weather and Marine Transportation Systems programs and is the archive for worldwide tsunami data. NGDC currently services 3.2 million requests per year and this number is expected to more than double by the year 2011.

How does this Center work? NGDC builds and maintains long-term archives of scientific data, with a special emphasis on scientific stewardship of data acquired by NOAA observing systems. NGDC is a critical part of the national and international scientific infrastructure and at the same time provides public domain data to commercial, military, and public users. NGDC works closely with contributors of scientific data to prepare documented, reliable data sets. Currently there are more than 400 digital and analog databases. NGDC's archive is currently 165 terabytes and will grow to over 1,200 terabytes by the year 2012.


NGDC continually develops advanced data management techniques that reflect the changing world of geophysics. Using the latest computing and Internet technology, NGDC will continue to develop new systems to provide easy access to digital data and products and develop an innovative decision support system. This will improve our ability to predict the impact of environmental factors on the economic health of our national infrastructure. Innovative technology will transform the world's largest collection of seafloor data into an integrated marine database, develop the first global database describing the long-term climate of the space environment, and provide the World Magnetic Model to the national and international scientific community.

Who are the Center’s Partners and Customers? NGDC works closely with Federal and state governments, industry, academia, and foreign contributors to build the data resources required to monitor, understand, and predict the environment. NGDC data and products support industry, education, and public requirements for environmental information. NGDC also hosts three International Council for Science (ICSU) World Data Centers.

For additional information: www.ngdc.noaa.gov and www.nesdis.noaa.gov
The National Oceanographic Data Center (part of the Oceanographic Institute of Ecuador), maintains a 50-year data base of continuous oceanic, coastal and atmospheric data collections based on oceanographic cruises, fixed meteorological stations along the Ecuadorian Coast and Galapagos Islands, in addition of monthly observations in two points in the coast and many in-situ samples (physical, biological, chemical, geological).

All the collected information has been organized in a computer system based on a ORACLE relational database named SIHO (Hydrographic and Oceanographic Information System) which was created based on the international standards of data management published by IODE (International Oceanographic Data Exchange) according to data types and units. The metadata was made using the DIF format (Directory Interchange Format) from IODE.

At the moment, all the collected data have a limited access, it is only delivered for research purpose to institutions, universities etc. Data products may be accessed by the web page www.inocar.mil.ec.

An important point is that Ecuador, represented by the National Oceanographic Data Center, has been working in the ODINCARSA project (IODE Project), whose mission is to strengthen Ocean Data and Marine Information Management capacity in the Caribbean and Latin America in order to contribute to Ocean Sciences, Operational Oceanography development and Integrated Coastal management activities at the regional level. Through this project, the Latin America region has been access to courses and workshops in oceanographic data management and other activities that have given a great contribution to the knowledge in oceanographic data management and have bring up a better communication between the data centers in the region.
The NEPTUNE Canada Cabled Observatory Data

The NEPTUNE Canada and VENUS projects represent a total of over $90M to establish innovative regional cabled observatory systems with hundreds of sub-sea scientific instruments in the deep sea off the coast of British Columbia. The Canadian contribution to what is bi-national US-Canada project is made possible by funding from the Canadian Foundation for Innovation and the BC Knowledge Development Fund. The University of Victoria hosts the two projects. The goal is to utilise the recent advancements in fibre optic communication technology to enable scientists on shore to probe the mysteries of the ocean and allow the world to share in the incredible discoveries. The development of powered, fibre-optic cabled observatories covering a wide range of marine environments will allow, for the first time, real-time observations that will establish a comprehensive interactive database and decadal time-series.

VENUS represents two cables of respectively 3 and 50 km with a total of three instrumented nodes. NEPTUNE Canada, to be installed partly in 2007 and 2008, represents a cable loop 800 km long, providing power and bandwidth to five instrumented nodes. The nodes will be central aggregation points for more nodes and junction boxes situated in the immediate vicinity (a few km) of the principal node. A total of about 150 instruments will be hosted and provide data from perhaps 600 different sensors.

To address a variety of scientific interests spanning many disciplines from geology to biology, the sensors will be of various types, measuring physical and chemical properties of the water and the seafloor. Scalar sensors will cover: temperature, pressure, salinity, diluted oxygen or nitrogen, but also transmissivity, salinity, conductivity etc. Multi-dimensional data structures will also be returned whereby images, velocity vectors etc. will be made available. Finally streams of data from hydrophones, video camera, seismometers etc will also be collected. All data types will be made available for immediate consumption by event detection or on-line visualization systems and be recorded for later analysis.

The NEPTUNE and VENUS observatories have decided to adopt open data policies. This means that the data are made available to all without restriction. To cover the rare cases where some data may have to be protected for a few weeks, the data policy foresees the ability to limit access to some well defined collection, after approval by the Science Advisory Committees of the respective observatories.

Access to the data by scientists and the public will be through web pages (search pages) and web services. Data format delivered so far include CSV, Matlab. NetCDF is in the planning. We are waiting for requirements from the community in this regard.
NIED Data Facilities

1. Data type(s) that are managed within your system

- High-sensitivity seismograph data: continuous recording
- Hypocentral catalog in and around Japan determined by JMA
- Broadband seismograph data: continuous recording
- Velocity type strong motion seismograph data: continuous recording
- Strong motion accelerograph data: event triggered data

2. Scope/mandate of the system (e.g. programmatic, spatial, or thematic scope)

After the disastrous 1995 Hyogoken-nanbu (Kobe) Earthquake, the Headquarters for Earthquake Research Promotion was established by the Japanese government and the “Fundamental Survey and Observation for Earthquake Research” project was set up in 1997. The purposes of this project are to estimate earthquake occurrence and evaluate strong ground motion based on the understanding of earthquake phenomena and the generating mechanisms of strong ground motion. Earthquake observations were selected as one of high-priority tasks.

3. Data policies (e.g. access restrictions)

We request all users to make registration to access NIED data. Please visit our home page about registration (http://www.hinet.bosai.go.jp/register/ENGLISH/). If users want to continue using their own account, they have to update them yearly. At the timing of renewal or at any time, users should inform us of their results which they have obtained by using our data set, because we have to evaluate the efficiency of our seismic observation data.

Second redistribution of our data is prohibited.

If you want to use data provided by national university, you have to contact a person who has responsible for each university network. Please see HARVEST home page in ERI, Tokyo University (Japanese webpage only); http://www.eic.eri.u-tokyo.ac.jp/harvest/

4. Access interfaces & tools (including web services, APIs, etc.)

- High-sensitivity seismograph data and Hypocentral catalog
  Visit Hi-net website (http://www.hinet.bosai.go.jp/; Japanese webpage only). Data within last 2 weeks are available by way of Internet web browser. If you want to use older data, send e-mail to hinet-admin@bosai.go.jp.
- Broadband seismograph and velocity-type strong motion seismograph data
  See F-net website (http://www.fnet.bosai.go.jp/freesia/index.html). All data are accessible on the Internet.
- Strong motion accelerograph data
  See the portal site of our strong motion seismograph networks. All data are available on the Internet. (http://www.kyoshin.bosai.go.jp/index_en.html)

5. Metadata standards, vocabularies

- High-sensitivity seismograph data
  We use original format called WIN32, but it is easily convert to ASCII, and SAC by using "WIN32tools" which is delivered by Hi-net web site.
- Broadband seismograph data
  We provide all data as WIN32, miniSEED, SEED, SAC, and ASCII.
- Strong Motion seismograph data
  We adopt our original format called K-NET format. Detailed explanation about the format is written in our web page. We deliver the utility program to operate K-NET format data.
The Ocean Data Bank (ODB) was established in 1985, and was subsequently merged into National Center for Ocean Research (NCOR) that was funded by the National Science Council (NSC) of Taiwan in 1997. Currently, there are mainly four different types of data available in ODB: shipboard ADCP data, hydrographic data, underway bathymetry data and seismic reflection data; as well as some minor heat flow measurements, underway gravity and magnetic data. The Sb-ADCP and CTD data covers those collected in the last decade and are indexed spatially. Composite current velocity at different depth have been compiled. More than 2000 cruises underway bathymetry data have been archived and processed since 1989. 500m x 500m gridded digital bathymetric model (DBM) dataset have been compiled. There are also more than 30 R/V OR1 cruises and 3 international cooperation cruises (1990 MW9006, 1995 EW9509 and 1996 ACT) multi-channel reflection seismic data that have been archived and transformed to SEGY format and stored in CD/DVD discs. There are limited access to these data for the purpose of academic researches. There is also a web link: http://www.ncor.ntu.edu.tw/ODBS/. Unfortunately, the English version of several essential descriptions are still in progress.
The OceanLife Project
The OceanLife project is a testbed collaboration between Rutgers University Marine Lab and the San Diego Supercomputer Center (SDSC) showcasing data mediation technology developed at SDSC. It is designed to assist marine scientists integrating biological and physical oceanographic data. Two of the larger datasets that it deals with are the World Ocean Atlas, containing physical oceanographic properties at one-degree resolution and many depths, and the Ocean Biogeographic Information System (OBIS), containing georeferenced observations of species. These species observations number in the millions.

Data Mediation
One of the primary foci of the Data and Knowledge Systems group at the San Diego Supercomputer Center is data mediation. In particular, we’ve developed a data mediation system handling many different types of relational sources, along with geospatial data. Vector data is considered an extended data type and most OGC functions on geometries are internally supported. This mediator system has been used primarily to integrate global physical oceanography data with biological oceanography data at as part of the OceanLife project. A query to the system may come in the form: given a range of dissolved oxygen and temperature, what species have been observed in these locations? Or: given a set of species’ observations, in what ranges of physical oceanography parameters are they found? The mediator system also features a prolog-like query language, XML API, and a web service provider which accepts queries. The web service component is very useful for integrating the mediator with existing tools such as workflow engines. Typically, however, an application specific user interface is developed using the XML API.

User Interface
A constant challenge in the GIS world is delivering massive amounts of geospatial data to a web client in a usable fashion. We’ve attempted to accomplish this with a combination of UMN MapServer and Google Maps. As users modify query parameters, the images rendered by MapServer (a combination of vector and raster) are updated and displayed as overlays on the base Google Map. Through MapServer, many different types of geospatial data can be effectively streamed onto a Google Map using WMS. In the OceanLife project these formats include: WMS, OPeNDAP, GeoTIFF, ESRI Shapfile, Oracle Spatial, and PostGIS. Because the geospatial data is not rendered by Google Maps, but is instead passed as a georeferenced image, many more features can be displayed. The application logic, including this mapping interface, is currently implemented using JSP.

Future Research
Currently, our research group is extending our mediation technology to support and query ontological data. These extensions cut across many projects, and in particular will be used to add habitat data stored as an OWL model into the OceanLife system.

Contact
Chris Condit: condit@sdsc.edu
The Ridge 2000 and MARGINS Data Portals of the Marine Geoscience Data System

Suzanne Carbotte, William B.F. Ryan, Robert Arko, Vicki Ferrini, Andrew Goodwillie, Andrew Melkonian, Suzanne O’Hara, Rose Anne Weissel.

Overview: Dedicated data portals for the Ridge2000 and MARGINS programs are provided through the Marine Geoscience Data System (MGDS, www.marine-geo.org), an integrated data system, providing open public access to a wide variety of marine geoscience data collected throughout the global ocean. System components include a cruise metadata catalog, digital data repository and a dynamic multi-resolution gridded synthesis of seafloor bathymetry data integrated with global land topography. Cruise metadata, including basic expedition information, data inventories and acquisition metadata reside in an open-source relational database (PostgreSQL®). FGDC standard metadata are created to enable exchange with the World Data Center for Marine Geophysics at the National Geophysical Data Center (NGDC). For users wishing to search the database for particular expeditions of interest, data discovery is via a text-based web search interface (Data Link) with options to search on key parameters including geographic range, ports, chief scientist and data types (www.marine-geo.org/link). The primary data visualization tool of the MGDS is GeoMapApp, a virtual map viewer coded in Java™, which permits dynamic exploration of a wide range of maps, grids, images, profiles, and tabular data with capabilities for customized map generation.

Data Holdings: A diverse and multi-disciplinary range of data are hosted through the MGDS ranging from temperature probe data, biological species compilations, MAPR and CTD data, to side-scan sonar, photographic transects, ultra-high resolution bathymetry data, as well as trackline geophysical and seismic data. Derived data products are also served including microseismicity catalogs, images and visualization scenes, magnetic and gravity compilations, and grids of seismic layer thickness.

Gridded Global Topography Synthesis: The primary data synthesis product of the MGDS is a dynamic global synthesis of ocean bathymetry derived from publicly available multibeam bathymetry data. The synthesis is maintained as a multi-resolution gridded global digital elevation model (DEM) to ~100 m spatial resolution and is merged with land topography data from the NASA Space Shuttle Radar Topography Mission. The MGDS provides synthesis of expedition based multibeam datasets at their full spatial resolution data for non-specialist use by maintaining a continually updated dynamic gridded global compilation. The global DEM can be accessed through GeoMapApp, and via a Web Map Service.

Data Submission: Digital metadata forms are provided for documenting data acquisition and sampling activities. All investigators funded by US MARGINS and Ridge2000 are required to submit complete expedition documentation and field data to the MGDS within time frames set by program data policies. The MGDS team works with scientists to ensure adequate expedition documentation and basic quality assurance of data.
provided. All data contributed to the data system are incorporated with access restrictions until explicit permission for data release is provided by the investigator, normally under the timetable specified in funding agency and program data sharing requirements.

**Web Services and Interoperability:** MGDS is working to make map imagery and underlying data sets available via OGC-compliant services (www.marine-geo.org/ws). Maps created from the global DEM are available as an OGC [Web Map Service](http://www.marine-geo.org/ws) (WMS) and can be accessed by any OGC-enabled client. An OGC Web Feature Service (WFS) provides direct access to station and sample data. MGDS is working to develop data interoperability and direct links to other repositories including NGDC, IRIS, UNAVCO, under the GeoWS project ([www.geows.org](http://www.geows.org)). An [OAI-PMH Provider/Harvester](http://www.geows.org) currently serves metadata records for multibeam surveys that conform to the NOAA Metadata Manager and Repository (NMMR) schema. Our entire suite of collections and vocabularies is available via [getCollection/getVocabulary](http://www.geows.org) REST-type services.
SeaDataNet: a Pan-European infrastructure for ocean and marine data management

Dick, D. M.A., Schaap, MARIS, The Netherlands, dick@maris.nl, technical coordinator
SeaDataNet, on behalf of the SeaDataNet consortium (see www.seadatanet.org)

Introduction:
SeaDataNet (2006 – 2010) is a major Pan-European and EU-funded project, undertaken by the National Oceanographic Data Centres (NODCs), marine information services of major national institutes, from nearby all coastal states bordering the European seas. The consortium comprises 49 partners of major oceanographic institutes of 35 participating countries, all riparian to European seas, acting as National Oceanographic Data Centres (NODC), Satellite Data Centres (SDC), two expert modeling centres and three international bodies. The NODCs and SDCs are highly skilled and actively engaged in data management for several decades and have the basic capabilities and facilities for data quality control, perennial safeguarding, retrieval and dissemination to both intermediate users (value adding organisations) and end-users.

SeaDataNet continues and expands previous initiatives of the consortium, in particular Sea-Search (2002-2006) (www.sea-search.net) and several distributed data management structures developed during MAST and the following EU marine environment projects. The recent developments made during Sea-Search were focused on metadata. This is continued and optimised in SeaDataNet, but moreover focus will be given to providing transparent and integrated data access by interconnecting the data centers.

Data types that are managed within SeaDataNet infrastructure:

These comprise in-situ and remote sensing data, metadata and data products. Disciplines include physical-oceanography, marine chemistry, marine geology, bathymetry, marine biology. These result from many national monitoring programmes and from research cruises. The first category has a focus on European seas; the latter category spans the global oceans and European seas.

Scope / mandate of SeaDataNet:

SeaDataNet is an infrastructure, which interconnects the National Oceanographic Data Centers and marine data focal points from 35 countries around European seas. These data centers are part of major and leading marine research institutes in these countries and have a long-term perspective. This implicates that the SeaDataNet infrastructure also has a long-term perspective as an operational and well embedded network of centers and systems. The SeaDataNet project has a five year funding from the EU and is used to develop common protocols, standards and software modules and to develop and populate joint metadatabases and data access portals. SeaDataNet is a successor project to the Sea-Search project, which run from 2002 – early 2006, and also received EU funding. Generally speaking the EU funding is used for concertation among the partners, while the actual implementation of standards etc is done by each partner at its own center. Further support by EU in the future is foreseen for further developments, horizontally and vertically, while the future of the centers itself is guaranteed at national level.

Data policies:
SeaDataNet has defined a common SeaDataNet Data policy, which is now under consultation with all partners. Its basis is striving for free, but controlled access to all data. Control includes, that users will have to register and agree with a common SeaDataNet licence for data access. Metadata browsing is in the public domain and requires no registration. At center level there might be local conditions to data access, which have to be respected. The SeaDataNet infrastructure will have a mechanism for searching interesting data sets (discovery services), ordering / requesting access to data sets, dealing with the local conditions (authentication and access services) and downloading and viewing services.

**Access interfaces & tools**

The SeaDataNet infrastructure will have a mechanism for searching interesting data sets (discovery services), ordering / requesting access to data sets, dealing with the local conditions (authentication and access services) and downloading and viewing services. The SeaDataNet infrastructure is built in 3 phases:

**Version 0 = present** : Users can search and browse in common online metadatabases including geographical interface and then connect or submit their requests for data access to individual user interfaces of partner data centres. And users can apply the ODV package for local analysis and visualisation of downloaded data. Common vocabulary web services.

**Version 1: from February 2008 onwards**: Users can search and browse in common metadatabases including geographical interface (OGC compliant) and then request access to data by downloading services via a common shopping mechanism annex download manager. The central tool will arrange that requests are distributed to the data centers and that the data sets are provided for downloading by a common interface and in common formats. The metadatabases will also be available as Web services for feeding local and regional portals. All metadata are compliant to the ISO 19115 metadata standard. The ODV package has a seamless connection to SeaDataNet output formats.

**Version 2: from 2010 onwards**: Virtualisation of data access and further viewing services.

**Metadata standards and vocabularies:**

Within SeaDataNet a number of Pan-European metadatabases are kept up-to-date by the partners:

- EDMED: Marine Environment Data sets dispersed in the scientific laboratories;
- ROSCOP/CSR Cruises Summary Reports;
- EDIOS : Initial Observing Systems;
- EDMERP : Marine Environment Research Projects;
- EDMO: Marine Organisations
- CDI: Common Data Index

These directories, available for consultation on the SeaDataNet website [http://www.seadatanet.org](http://www.seadatanet.org), and represent a key tool for data search and retrieval. The directories are compliant to ISO 19115 metadata standard and depending on the metadatabase a number of entry tools are available (offline entry, online entry, XML exchange).

For harmonisation between the metadatabases a central Common Vocabularies web service has been set up by partner BODC, including an international content governance mechanism. Next to these also a common directory of marine organisations (EDMO) has been set up as Web service by partner MARIS and this is maintained by online CMS by the centres per country.
The SeaDOG archive and its integration into RODIN

The SeaDOG (Stewardship of Deep Ocean Geophysical) database continues to grow with data (e.g. Bathymetry, Sidescan, Magnetics, Gravity, Seismic and others). Data contained in the database is usually created from cruises or project data and is both in raw and processed states. The database contains metadata information from over 275 cruises and has indexed over 63000 datafiles. This equates to over 1.8 Tb of data. Much of the metadata acquisition is automatic though it still requires the PIs to be encouraged to actually present their data for archive.

This project was initiated in 1998. After the initial 3 years funding this indexed archiving was continued as it was found to be the only viable way to properly manage NOCS geophysical data without spending vast sums on database integrations and quality control. Initially the system held the actual data offline for security and as the data sizes were too large. Fortunately the data can easily be archived on-line (or near on-line).

RODIN (Repository of Oceanographic Data and Information at NOCS) has now been created to act as a portal for the SeaDOG data and to aid integration into the UK NERC data grid (NDG). The system consists of four main parts: the user and system interfaces, the data catalogue and related modules, the data archive and related modules and the modules interfacing to the NDG.

- The data catalogue contains metadata, information about location of data archived, permissions on data and metadata, information about users, privileges and operations concluded or to be performed on the data and finally a history of accesses. Access to data and metadata is regulated by an Authorization Module. An Administrative Tasks Module is in charge of extracting and translating metadata into formats (DIF, MOLES, ISO19115, etc.) suitable for communication with external services.
- The data archive consists of a robotic tape system managed by the Tivoli Storage Manager (TSM). The Archive/Retrieve Module checks routinely the Data Catalogue for new archive/retrieve requests, calls the TSM and stores information about the results of the operations back into the Data Catalogue.
- The System Interface allows different systems to connect to RODIN. The Authentication Module checks that users are registered with the system and assigns them the privileges granted by their user profile. A web GUI and a Command Line Interface are available. The CLI accesses the system through a web service wrapper. Other modules, included in the NDG interface, allow metadata harvesting by means of an OAI (Open Archive Initiative) infrastructure and metadata browsing based on the MOLES (Metadata Objects for Links in Environmental Sciences, a metadata format developed by NDG) repository.

The stewardship of geophysical data has three objectives:
- Retrieve, collate, validate, curate and preserve existing data as a valuable resource so that it is not lost. Thus it becomes available to be used and exploited by the scientific community and other end-users, including commercial interests.
- Supply data ‘freely’ to researchers in the academic community. This would be done as a service at minimal cost, either totally free of charge, or with a basic handling costs (e.g. media cost) or at discounted prices according to circumstances.
- Provide adequate facilities for the stewardship, validation and dissemination of the data, together with maintaining standards of the data.

Tim Le Bas, Fabio Venuti (NOCS, UK) – May 2007
The Integrated Ocean Drilling Program (IODP) is managing data from three scientific ocean drilling platforms, operated by the United States, Japan and European science operators consortium, each using different data capture and management systems. To meet the need of integrating data access from distributed sources, IODP is developing the Scientific Earth Drilling Information Service (SEDIS¹), that will be a unique entry point for finding all IODP data, as well as legacy data from previous scientific ocean drilling programs (DSDP², ODP³). SEDIS is not a data repository but the gateway to distributed data repositories.

The types of data are extremely varied from typical scientific measurements (e.g. at a location and/or versus depth), images, scanning and descriptive data. SEDIS will handle the search, discovery and access to all scientific data captured during and post drilling expeditions including all publications related to the program regardless of their storage location. All data is publicly available after the moratorium period (usually one year). The development of SEDIS is taken out in three distinct phases:

- **Phase I:** Dataset discovery using a metadata catalog
- **Phase II:** Program publication and citation full text search engine
- **Phase III:** Advanced data access and visualization tools

SEDIS Phase I will allow the user to discover all type of datasets for all drilled holes, via web based interface and web services. Each dataset provided by the contributors will be described using the ISO 19115 metadata standard and ISO 19139 XML implementation. It will also provide links to the datasets themselves, independently of acquisition, storage location or archival status. Each dataset is expected to be assigned a Scientific and Technical Data Digital Object Identifier (STD-DOI⁴) so that external publications and databases can easily link to the metadata. SEDIS Phase I is currently under development by MARUM, Bremen, Germany, and will be launched in the last quarter of 2007.

Phase II is intended to tie together the program data and publications. This includes the capability to harvest, store and manage metadata (using Dublin Core) about current and legacy program publications in various electronic formats (HTML, PDF, etc.), index the content and provide full text searches via web interfaces, web services and Google Scholar. Publication search results will also be linked to related datasets. As for the datasets, each publication will also be assigned a Digital Object Identifier (DOI).

Finally SEDIS Phase III will include advanced tools enabling users to search, extract, and combine the data itself as well as to integrate advanced data visualization and mapping tools as they become available. The data will still come from the different contributors and is expected to be formatted and streamed using a common XML format⁵. The goal is to enable the user to extract any part or combination of data from any drilling expedition, and to provide methods to process and/or transform these data with online viewing, mapping, and manipulation tools from a single entry point.

---

¹ SEDIS: [http://sedis.iodp.org](http://sedis.iodp.org)
² DSDP: Deep Sea Drilling Program ([http://www.deepseadrilling.org](http://www.deepseadrilling.org))
³ ODP: Ocean Drilling Program ([http://www.odplegacy.org](http://www.odplegacy.org))
⁴ STD-DOI: [http://www.std-doi.de/](http://www.std-doi.de/)
⁵ Observations and Measurements: [http://www.opengeospatial.org/standards/bp](http://www.opengeospatial.org/standards/bp)
Data Facility Description

SIOExplorer Digital Library
Geological Data Center
Scripps Institution of Oceanography, UCSD
La Jolla, CA 92093-0220 USA
http://SIOExplorer.ucsd.edu

Data types
Multibeam bathymetry swath, grid and visualization files
Sidescan
Navigation
Underway geophysics with magnetics, gravity and depth data and profiles
Seismic reflection and profiler
Current profiler ADCP, HDSS
Meteorology
CTD, XBT
Cruise, QC, narrative and troubleshooting reports
Dredged rock and core sample descriptions
Historic photographs
Resources for K-12 teachers, primarily on plate tectonics
Seamount studies

Scope and mandate
Shipboard data, images and documents from SIO expeditions since 1950. 753 cruise legs online, including 300 multibeam cruises. Concentrating on Pacific and Indian Ocean, some Atlantic and Mediterranean cruises. Collection growing to 1000 cruises this year. Extension to WHOI holdings work in progress, cruises, Alvin dives and ROV lowerings, primarily in Atlantic.

Data policies
Approximately 90% available for immediate public release. Objects on academic proprietary hold require permission from cruise leg chief scientist.

Access interfaces
Web query and interactive geographic Java interfaces

Metadata standards and vocabularies
SIO/SDSC metadata template file approach
Controlled vocabulary with authorized values and dictionary for metadata parameters
Data Facility Description

Site Survey Data Bank
Geological Data Center
Scripps Institution of Oceanography, UCSD
La Jolla, CA 92093-0220 USA
http://ssdb.ioldp.org

Data types
Active deformation, backscatter, bathymetry, documents, electromagnetics, environmental conditions, fluid flux, geology, gravity, heat flow, location maps, logging, magnetics, navigation, permits, sample data, seafloor imagery, seismic (3D, multichannel, single channel, refraction, profiler, velocity models, seismicity), visualization, water column

Scope and mandate
Currently 4800 digital objects in a digital library to support the review of IODP ocean drilling proposals, worldwide.

Data policies
All users must register. Approximately 90% available for public release. Requests for held objects managed by IODP-MI.

Access interfaces
Web query and interactive geographic Java interfaces.

Metadata standards and vocabularies
SIO/SDSC metadata template file approach
Controlled vocabulary with authorized values and dictionary for metadata parameters
SNAP (Seismic data Network Access Point)
Paolo Diviacco
(Istituto Nazionale di Oceanografia di Geofisica Sperimentale, OGS)

SNAP is a web content manager devoted to Seismic data sharing and dissemination and aimed at fostering the collaborative attitude of Scientific/Academic institutions.

1. **Data type(s) that are managed within your system**
   - Seismic data: SGY files
   - Scanned seismic sections: tiff files
   - Unstructured files: txt, pdf, jpg, gif (any data type, that linked to a specific seismic line report acquisition, processing or interpretation steps).
   - Positioning: ukooa, svg
   - It is anticipated that other data types will be soon available as: well logs, multi-beam, oceanographic data etc.

2. **Scope/mandate of the system (e.g. programmatic, spatial, or thematic scope)**
   The system is a framework that can be used within several initiatives with different goals. In the case of the Antarctic Seismic Data Library System (SDLS) public marine seismic data from Antarctica is fully and immediately available to anyone. In other cases, as the EU project CO2GeoNet, the system serves a closed community as a reference although distributed data access facility. It can be at the same time spatial and thematic while mainly used to trigger new collaborations with external data seekers.

3. **Data policies (e.g. access restrictions)**
   Users need to be registered (a guest user is provided that can access to only partial functionalities for demo purpose) After login they are granted the lowest data access privileges and can visualize files only with protective resolution or watermarking. After browsing the data available in a single server or in the central aggregator of a distributed system, once they found the data they are interested in, users need to negotiate with the data owner one of the data access policies available: from viewing only the data, to collaborative data access trough a Computer Supported Collaborative Work (CSCW) toolkit (groupware) as the integrated solution developed within OGS called COLLA.

4. **Access interfaces & tools (including web services, APIs, etc.)**
   - **Seismic data**: depending on users privileges, can be interactively web-viewed (zoom, pan, gain) within a server side paradigm, or, used within a Computer Supported Collaborative Work (CSCW) tool, or if needed downloaded.
   - **Scanned seismic sections**: tiff files, converted to jpg files, can be web-viewed and, depending on users privileges, presented with or without a “protective” low resolution, watermarking, or CSCW / download possibility.
   - **Unstructured files**: depending on users privileges can be viewed, used within a CSCW or downloaded.
   - **Positioning**: is sent client side as an SVG (Scalable Vector Graphics) file and there rendered, unloading the server, reducing the network traffic and obtaining light-speed map zooming and panning without pixellation.

5. **Metadata**
   Metadata is directly extracted from the actual data during uploading, in order to reduce at minimum, when not completely avoiding, data editing. In the case of seismic data and positioning this is possible since the data itself is produced by acquisition and processing software that always refer to community driven “traditional” data models. Interfaces to ISO and PPDM standards are on the way, in the view of maintaining minimal the editing needs while ensuring compatibility.
Following the metadata standards and conventions as delivered by the different data holders.
WDC-MARE / PANGAEA – Data publication and information exchange

Major challenges

Fostering **data management as an integral part of science.** Data driven science needs early involvement of data centers in science project planning, product & services specification, and conception and implementation of observing and measurering facilities. To be more than just an archive data centers must take an active role in all relevant programs, projects, and institutional plannings.

Data products compiled from many sources need high sophisticated data networks. They are the necessary prerequisite for complex and large scale scientific approaches. Implementation of **Spatial Data Infrastructures (SDI)** on a global scale is a major challenge for all institutions involved in the production, archiving, and distribution of scientific data.

The poor availability of data is a major structural problem in empirical sciences. The exponential increase of scientific data and information requires new and more concise forms of publications. Reliable and high quality **peer reviewed data publications** will compensate for the current deficiencies in scientific publishing.

Technical Organisation

**WDC-MARE** uses the information system **PANGAEA®** as operating platform. The basic technical structure corresponds to a three tiered client/server architecture with a number of clients and middleware components controlling the information flow and quality. On the server side a relational database management system (RDBMS) is used for information storage. To ensure fast data access the data are mirrored in data warehouse which is also used as interface to the German GRID community. All interfaces to the information system are based on web services including a simple map supported (Google Earth & UMN) search engine.

With its comprehensive user interfaces and the built in functionality for import, export, and maintenance of information **PANGAEA®** is a highly efficient system for scientific data management and data publication

Metadata

The challenge of managing the heterogeneous and dynamic data of environmental and geosciences was met in **PANGAEA®** through a flexible data model which reflects the information processing steps in the earth science fields and can handle any related analytical data. Metadata are consistently stored in a relational database and can be served as ISO115, DIF, FGDC, or DC compliant catalogues.

The data description of any data set includes the principle investigators (PI) name and email for contact, parameters, used methods as well as spatial and temporal coverage.

Data content

The WDC-MARE manages data related to Global Change in the fields of environmental oceanography, marine geosciences, and marine biology. It focuses on georeferenced data (numeric, text, and any kind of binary objects)

Except a few data from ongoing projects which are under moratorium all data are freely available and can be used by referencing the related publication or the data set citation.
Major activities

Data management for scientific projects.
Since 1996 PANGAEA® supplies data management services on an international level. Until 2007 PANGAEA® was and is a partner in more than 60 European to international projects covering all fields of environmental sciences.

Networking & data portals
During the last years the PANGAEA® group got engaged in a number of projects supporting spatial data infrastructures (SDI). PANGAEA is the lead partner for the implementation of portals and infrastructures for the International Ocean Drilling Project (IODP), the NoE EUR-OCEANS and IPs CarboOcean and HERMES. The group has substantiated knowledge and practical experience on international standards as ISO19xxx family of standards, OGC standards - in particular the Sensor Web Enablement (SWE) concept -, Open Archives Initiative protocol (OAI-PMH), GRID technologies, and data portals.

WDC-MARE / PANGAEA is operated by the

Alfred Wegener Institute for Polar and Marine Research (AWI), Bremerhaven, member of the Helmholtz Association of National Research Centres, funded by the Federal Ministry of Education and Research and the

Center for Marine Environmental Sciences (marum) at the University of Bremen with support of the Department of Geosciences and the research center ocean margins (rcom), funded by the German Research Foundation