Report of the iSSP Data Bank Working Group

5 March 2003

Introduction

The anticipated transition from the JOIDES Resolution-based ODP to a more complex, multi-platform IODP has sparked a great deal of interest in how the data necessary to support scientific drilling must evolve and how those data should be stored and accessed by the scientific community. The former ODP Site Survey Panel, along with the ODP Data Bank, had occasional discussions about the future, and the interim Site Survey Panel (iSSP) has increasingly been grappling with these issues over the past year. Recognizing the need to make more rapid progress the iODP Data Bank Working Group (DBWG) was formed as a subset of the iSSP panel and liaisons from iPPSP and iSciMP. This group has met partially, or in full, three times, with the most recent meeting occurring before the iSSP meeting February 2003 in Bologna, Italy. This report presents the results of our meetings with the idea that our recommendations will help shape the request for proposal for the successor IODP Data Bank. Following our meeting structure and as a direct response to the iPC consensus statement 2-5, the report is organized to address points 1-4 in that statement. As a preface, we include a Data Bank mission statement, which will hopefully demonstrate our intended purpose and philosophy.

Participants in DBWG meetings (members of iSSP unless otherwise indicated):
David Divins (iSciMP), Andre Droxler, Nobu Eguchi (iSAS), Kirk McIntosh, Daniel Quoidbach, Craig Shipp (iPPSP), Tetsuro Tsuru, and Joel Watkins (iPPSP/iSSP liaison)

Data Bank Mission Statement:
The fundamental mission of the IODP Data Bank is to receive, catalog, and store data necessary to support science and safety for IODP drilling activities. An equally important mission of the Data Bank is to maintain a system to disseminate these data as effectively as possible to IODP panel members and to participants on the various drilling platforms. While the DBWG and iSSP recommend that access to these data be as broad as possible to encourage community involvement, the Data Bank must also control access to its contents, under the guidance of the IODP CMO, to protect the proprietary nature (at various levels from commercial to first rights of investigators) of many expected data sets.

iSSP DBWG Recommendations

1. Requirements for digital vs. analog data submissions:

We recommend that data submitted to the IODP Data Bank (DB) be in a digital form unless this is not possible for the proponent(s) to accomplish. This means that data including survey data, maps, and reports that are computer generated or processed should be submitted in a digital format (to be discussed below), and it also means that if only analog versions of the data exist, then these should be scanned by the proponents and submitted as digital data files. Only in the case that required data do not exist in a digital form **and** the proponents do not have facilities to scan the data would it be acceptable to submit paper copies to the DB. As this will inevitably take place, however, the DB must be equipped with computers and large format scanners.

Another important aspect of digital data submission is the question of what formats are most desirable and most useful to the DB and the community it serves. Specifically, there is a basic divide between digital **images** (e.g., PostScript, Tiff, JPEG, etc.), whose primary purpose is to be viewed as is, and digital **data** (e.g., gridded or tabulated data in ASCII or binary, seismic data in SEG-Y, etc.), which may be used to create images, may be numerically analyzed, or manipulated and then viewed. In part the recommended format will depend on the software/hardware infrastructure that is adopted by IODP and the DB. If possible, we recommend a GIS capable software system, which would allow the most versatile display of text, digital image files, and digital data files and would still be web-accessible for easy remote viewing. We discuss software capabilities further in #4 below, but these systems do exist and are used in the oil industry.

In our preferred configuration the DB would receive digital **data** files of geophysical survey data and map data (in gridded files) as well as digital **image** files of fully annotated survey data and maps presented at a scale that reveals details necessary for scientific evaluation. The data files would be loaded in a GIS-capable, web-accessible, software system (similar or related to Landmark or GeoQuest seismic interpretation packages), and the image data and text files would also be loaded into this system for viewing. The digital data files would be used for in-depth evaluations by panel members (e.g., SSP, PPSP, and SSEP), and by the actual drilling leg participants. The image files would be used more as a quick data reference but would be a significant improvement over the few, tiny images that are possible to fit in page-limited drilling proposals.

We suggest that data be submitted as a standardized “REPORT” type package of figures and information, possibly with the submission of the pre-proposal or with full proposal submission. The REPORT would include a specific set of maps (regional and site-focused), key seismic profiles, and other appropriate data. These data would be submitted as images (.ps, .pdf, etc.), and their main purposes would be to provide a quick reference of data quality and availability, allow for early stage review by SSP and PPSP, and they would be

available for other panels such as SSEPs, when better data displays are necessary to evaluate the proposed science. REPORTs would be web-accessible to panel members (with correct permission or password) and could also be included on CDs produced for the panel meetings. To get an idea of what we mean by the term REPORT, a preliminary template, based on U.S. Minerals Management Service (MMS) requirements, is provided below. Not all items below would be appropriate for all drilling environments/objectives, and not all required items would be expected to be submitted simultaneously or even in the early stages of the proposal review process:

1) **Location Map** (page-size). Location map should show approximate location of study area with proposed drill locations, possibly including seismic coverage and bathymetry.
2) **Contoured Bathymetry Map** (typically 1:12,000 scale). Contours are to be labeled in meters below sea surface and contour interval is to be such that seafloor shape is fairly portrayed without impairing easy use of the map. Typical contour intervals expected on the continental shelf range from 1 to 10 meters; 10-meter or greater intervals are typical on much of the continental slope.
3) **Geologic Features/Geohazards Map(s)** (typically 1:12,000 scale). Geologic features/geohazards maps are to show all detected natural and man-made features (including all unidentified magnetic anomalies) that could adversely affect the planned drilling operations
4) **Data Examples.** For example, one annotated data example for each survey tool operated could be included to demonstrate typical data quality and geologic conditions. Additional data examples shall be added to help illustrate and explain the interpretations and conclusions reached. The data examples (and maps) should be submitted in as .ps, .pdf, or other electronic format, and should be produced at a size that will show sufficient detail for evaluation (generally page size at minimum).
5) **Shallow Structure Map(s)** (typically 1:12,000 scale). At a minimum, one shallow structure map is required at sites where strata are not horizontal or not continuous over the study area (that is, either the shallow horizons are dipping, faulted, or have been locally eroded
6) **Isopach Map(s)** (typically 1:12,000 scale). At a minimum, an isopach (thickness) map is required on the continental shelf if inferred weak surficial strata overlie a potentially stronger substrate. This will help to indicate areas that may need further analysis to determine jack-up punch-through potential.

2. Allowable Data Formats

The DBWG recommends accepting data in formats as described in the table below. We have tried to account for all typical types of data that are required to support scientific drilling. We have assumed adoption of a GIS-type, web-enabled software system, which favors submission of digital data (e.g., x, y, z in ascii files) rather than images for map-type data sets. In addition to the actual data, metadata describing the data formats, acquisition and processing parameters, map projections, etc., will have to be submitted. Metadata forms to recover this information have been developed by the current ODP Data Bank and will be used for this purpose.

Data Type	Format	Media
Seismic Data - Hi-Resolution - Deep Penetration	**SEG-Y Files*** Paper Profile *DB will **not** process seismic data. Data must be submitted as stacks or migrations along with supporting metadata.	**DLT** **8MM** **CD-ROM** **DVD-ROM** **FTP** **IBM Cartridge Tape** Paper copies

Seismic Velocities - Time-Depth Curve - Check shots - Velocity Model - Stacking Velocity	**Image file for velocity model** ASCII Files (clearly annotated) Table of values	**E-mail** **FTP** **CD-ROM** Floppy Disk
Sub-bottom Profiler - chirp, parasound, etc. - 3.5 kHz	**Image File** SEG-Y Paper Profile	**DLT** **8MM** **CD-ROM** **DVD-ROM** **FTP** **IBM Cartridge Tape** Paper copies
Maps - Swath Bathymetry - Side-looking Sonar - Contour Map - Other	**Image File (PS, Tiff, etc)** **Document File (.pdf, .doc, etc.)** Paper Map	**E-mail** **FTP** **CD-ROM** Floppy Disk
Gridded Data - Magnetics - Gravity - Bathymetry	**Grid data file** ASCII XYZ file Image File	**E-mail** **FTP** **CD-ROM** Floppy Disk
Digital Images - Seabed Conditions	**Image File (PS, Tiff, etc)** **Document File (.pdf, .doc, etc.)**	**E-mail** **FTP** **CD-ROM** Floppy Disk
Heat Flow - Tables of values - Plots/graphs of values	**ASCII Table** **image file (PS, Tiff, etc)**	**E-mail** **FTP** **CD-ROM** Floppy Disk
Document Files - Core Descriptions - Ice Conditions - Current/Tide Data - Sample Descriptions	**RTF File** **PDF File** Word Document	**E-mail** **FTP** **CD-ROM** Floppy Disk
Log Data	**LAS format files**	**E-mail** **FTP** **CD-ROM** Floppy Disk
OBS Microseismicity	**ASCII File** (clearly annotated) Image File	**E-mail** **FTP** **CD-ROM** Floppy Disk
Navigation	**UKOOA** MGD77 ASCII File (clearly annotated) SEG-P1	**E-mail** **FTP** **CD-ROM** Floppy Disk
Video (e.g., seafloor images of target area)	**Digital video (mpeg?*)** Restricted to immediate drilling area. *This category requires further research to determine the	**DLT** **8MM** **CD-ROM** **DVD-ROM** **FTP**

	optimum formats.	

3. Mechanism and timing of communications with IODP panels and proponents.

The DBWG recommends continuing the current policy of early review unless it proves to be unworkable when proposal numbers increase. Currently iSSP gets pre-proposals that have been reviewed by SSEPs and a full proposal has been requested. The purpose of an early iSSP review is to give proponents a preliminary idea about what sort of data are likely to be required and allow them to start planning/proposing surveys. The iSSP plans to encourage proponents to submit data to the Data Bank as it becomes available rather than waiting until later stages of the review process. This will become particularly important for the more complex programs involving deeper drilling (riser and non-riser) in complicated structural settings. Early data submission and evaluation allows more time to acquire additional necessary data (i.e., to image the target) and helps to avoid data related drilling delays.

The iSSP suggested a possible further improvement of this system during the Bologna meeting (February 2003). A new working group (the MATRIX working group) is, among other things, investigating the possibility of creating an automated, preliminary review system, which would be made part of the proposal submission web page. This preliminary review would be based on proponent answers (multiple choice) to key questions about the tectonic environment, target depth, and sediment thickness at the proposed drilling sites. The results, i.e., expected data requirements, could be available instantly, although a disclaimer would indicate the preliminary nature of the results and the requirement of further human participation (iSSP panel) in the process.

4. Facilities, hardware, software, and personnel for Data Bank

As noted above in #1, the DBWG recommends that IODP adopt a GIS-capable, web-accessible, software system. These types of systems are available from vendors such as Schlumberger and enable groups to view seismic data, gridded map-type data, image files, and text documents remotely using only a standard web browser. This type of system is particularly attractive because it allows seismic users to zoom, change the display type (wiggle/no wiggle/variable density/etc.), the color palette, and even manipulate the gain level. Map and log data can also be accessed remotely and manipulated to create optimum views. Another attractive point of these systems is that they are mature and functioning and could be put in place with a minimum amount of delay. There is also flexible implementation of these systems such as an ASP (application service provider) model where the data storage, data server, and application server are all maintained by a commercial service company. A similar system can also be implemented within an IODP facility with the service company providing mainly the software packages and technical support for installation and maintenance.

Although we were unable to research the available products thoroughly at this point, a representative of Schlumberger, Shigemi Matsuda, made an informative presentation to the DBWG at our one-day meeting in Bologna. Mr. Matsuda is responsible for the operations at the National Data Repository in Japan, which services the Japan National Oil Corporation's worldwide operations. We want to make clear that we are not endorsing this proposed plan but use it here for informational purposes, i.e., what is available and approximately what it costs. We suggest that full demonstrations of functionality be required in the process to evaluate any IODP Data Bank proposals. This is in part to demonstrate how effectively the systems work and also to help identify those specific product offerings that are necessary to operate the Data Bank.

We include below information from Mr. Matsuda's presentation as an example of a possible Data Bank implementation. We note that Mr. Matsuda expressed interest in operating the IODP Data Bank, with the intention of co-locating it at the current Japan National Data Repository (NDR). The information is provided largely in a table format and describes the hardware, software, personnel, and approximate costs that may be required to operate the IODP Data Bank and provide broad but secure access to the data holdings via the internet (WWW). This information was gathered from a document Mr. Matsuda provided to the DBWG; the full contents of Mr. Matsuda's presentation are available from the DBWG.

Information Technology

Network, Security, and Internet Access

Schlumberger Network Solutions group provides Schlumberger Connectivity Center (SCC). Depending on the number and types of external connections that are needed, two or more firewalls are necessary. Customer connections and services are protected and isolated on separate firewalls. A Xylan switch is needed to provided the paths between the multiple firewalls. All external connections are terminated on a router outside the firewall.

SCC Cost

Item	Description	Cost Estimate
Firewall SW and HW	Internet Gateway: Firewall HW&SW Services Gateway: Firewall HW&SW Sinet Gateway: AN Router upgrade FM management tool: SVR router Others	$52,000- (lease possible)
SCC setup	SCC customer connectivity setup charge	$8,515-

Data Preparation and Data Loading

Digitizing

Those data provided on paper must be digitized before loading into Database Software.

- Seismic Profiles
- Paper Maps
- Reports

Seismic line and Shot point location, well locations, contour lines, culture data, etc., often required digitization in order to load them to database then upload on WEB based GIS-map.

Data Loading

Due to the fact that most of the data types (cultural, navigation, E-docs and maps, Gravity & Magnetics etc) can be handled in Finder (GeoQuest Data Management Software) and there will not be huge volumes of seismic bulk data, there is likely no need to have a dedicated seismic archive system. The main idea is to use Finder to manage all the data types mentioned in the data formatting guidelines even for seismic bulk data and well logs which can be handled as E-Doc and associated with seismic lines and well locations.

Software Required

Item	Description	Est. Cost
Digitizing MAP/LOG software	to digitize hard copy map and log	$ 18,000-
Graphic software	to edit graphics	$ 2,000-
Finder	GeoQuest Database software offers a broad data model and an array of visualization tools.	$ 134,055-
	Cost Estimate	$154,055-

Hardware Required

Sun Blade 2000 Workstation System	Finder database server	$ 25,800-

Color A0 Scanner	to scan and digitize big size map and long size logs with color
Desktop Scanner	to scan and hardcopy document and standard size maps
PC	to run scan and digitizing software
Cost Estimate	$30,000-

Facility

Data Storage Room

Storage space is needed to receive physical data, including the proposals and data from the proponents, plus data inherited from the ~100 currently active projects in ODP being stored at the Data Bank in Lamont. The types of data anticipated include the following:

Hardcopy reports, maps logs

- Hardcopy document
- Reports, Maps and Logs in both black & white or color.
- Electronic document

Logging data

- These formats will be acceptable: LIS, LAS, DLIS, SEGY (VSP only). Other formats may be loaded as documents

Seismic data

- Navigation data - Provided in digital UKOOA format
- Bulk trace data - Post-stack SEGY format data.

Work Space and Computer Room

Space required for Data Preparation and computer room where important database and web server machines can reside securely.

Software Required

ASSET management Software	to control and manage data reception, return, rental and duplication, etc..
Resource	Librarian/ Clerk
Cost Estimate	?

Hardware Required

STORAGE space	Room to store physical data
Data Container	For example cabinet
Workspace	Room to do data preparation and desk work
Computer Room	Dedicated computer room for servers
Cost Estimate	?

Human Resources

The following personnel to be considered as minimum to run New IODP Data Bank as described in this proposal.

	Qty	Role	Fulltime or Part time
Project Manager	1	Report progress of the project. Communicate with CMO and Panel of scientists and proponents	Fulltime preferred
System Engineer	1	Regular back up of the system, Firewall rule change, system upgrade	Part time OK but quick response required in case of emergency
Database Administrator and Data loader	1	Database and Web User Management, Data Backup regularly QC-data Digitizing and Data load to database	Part time OK
Librarian	1	Data reception and shipment	Part time OK

Cost Estimate for IODP Data Bank located at Japan NDR

	Initial Expenditure	Monthly Payment
Data@Work purchase and Set up	$ 150,000-	-
PC server for Data@Work	$ 5,000-	-
Firewall Software and Hardware	$ 52,000-	-
SLB Connectivity Center	$ 8,515-	$ 4,000-
Finder Data Management Software	$ 0-	-
Digitizing Job	-	As per NDR price Book
Data loading	-	As per NDR price Book
Finder server workstation	$ 25,800-	-
Data Bank Monthly Maintenance Charge	-	$ 16,800-
Total	**$ 241,315-**	**$ 20,800- /month**