Report of the iSciMP DatabaseWorking Group

2-3 June 2003

Introduction

As the Ocean Drilling Program comes to an end and a new era of ocean drilling begins with the Integrated Ocean Drilling Program (IODP) new opportunities to explore our Earth will arise. The "I" in IODP will present the most challenges especially for the data management and the integration of database services throughout the new program. The task of the interim Scientific Measurements Panel's (iSciMP) Database Working Group (DBWG, hereafter called the Group) was to present a possible model for database services, which the Group refers to as the IODP Information Services (IIS). The model comprises the management of the data collected onboard the various platforms (including downhole logging, site survey information), legacy data from DSDP and ODP, and "landborn" data, derived from post-cruise research and publications. The model includes the integration of those data and other IODP relevant information types into a common, program-wide IODP information system accessible by IODP researchers and the public. This report presents the results of a meeting held 2-3 June at which the Group discussed future IODP database and data management activities. We begin the report with a preliminary "Mission Statement", or "Mandate", for an envisioned IODP Information Services Center (ISC), which will play a key role in the successful function of the IIS. This is followed by a set of recommendations for the functions and structure of the proposed ISC, including expectations for each of the IODP Implementing Organizations (IOs), their relation to the ISC, and a number of database management issues.

The DBWG Report makes specific recommendations, however it does not specify exactly how the recommendations should be implemented. This is done intentionally. There are many possible configurations and designs that will include all of the Group's recommendations, but the Group felt it was not its charge to define the specifics. Rather, the Group would present concepts that it believes will make for a successful IODP database management structure.

The proposed model for database management for IODP is highly flexible. This report encompasses the data collected by the various operational platforms with respect to cores (e.g. data currently collected by ODP), ODP and DSDP legacy data, post-cruise data, publications information, downhole measurements, seismic images, engineering data, and much more. However, the system is versatile and should include links to the Site Survey Data Bank and downhole logging database

Participants in DBWG meeting (members of Working Group unless otherwise indicated):
Jennifer Anziano (JOI), David Becker, Michael Diepenbroek, David Divins, Colin Graham, Hisao Ito (iPC), Shin'ichi Kuramoto, Kate Moran, Saneatsu Saito, and Kyoma Takahashi.

IODP Information Services Center Mandate

The IODP Information Services Center provides for the ready access of all IODP data to IODP researchers, the international science community, industry, educators, media, and the public in a timely manner. This is achieved through the coordinated actions of the Center and the Implementing Organizations in the development and implementation of common program policies, standards, and effective mechanisms for the collection and distribution of IODP data.

Recommendations:

1) Structure of IODP Information Services

The Group recommends that an IODP Information Services Center (ISC) be established to provide database services within a distributed networked system and not within a centralized system. The system, termed the IODP Information Services, is composed of the database management activities of each of the IOs, a database of legacy data (DSDP and ODP, where these data will be maintained is not specified), and, at its heart, the Information Services Center operating directly under the IMI (Figure 1).

The primary functions of the ISC should include:

- *a **clearinghouse function** provided by ISC management, technical, and communications staff with appropriate network and computer infrastructure to provide integrated access to the program-wide information; and*
- *a **coordination function** provided by an assemblage of information services staff from each of the IOs as well as the ISC, site survey data bank services staff, and scientific drilling legacy data staff.*

Discussion:

The Group envisions two major challenges to the new ISC. First, providing integrated access to all IODP data, ODP legacy data and DSDP legacy data. Second, working with the information services staffs of the IOs and those of other data providers to ensure that data structure and access standards are in place and followed.

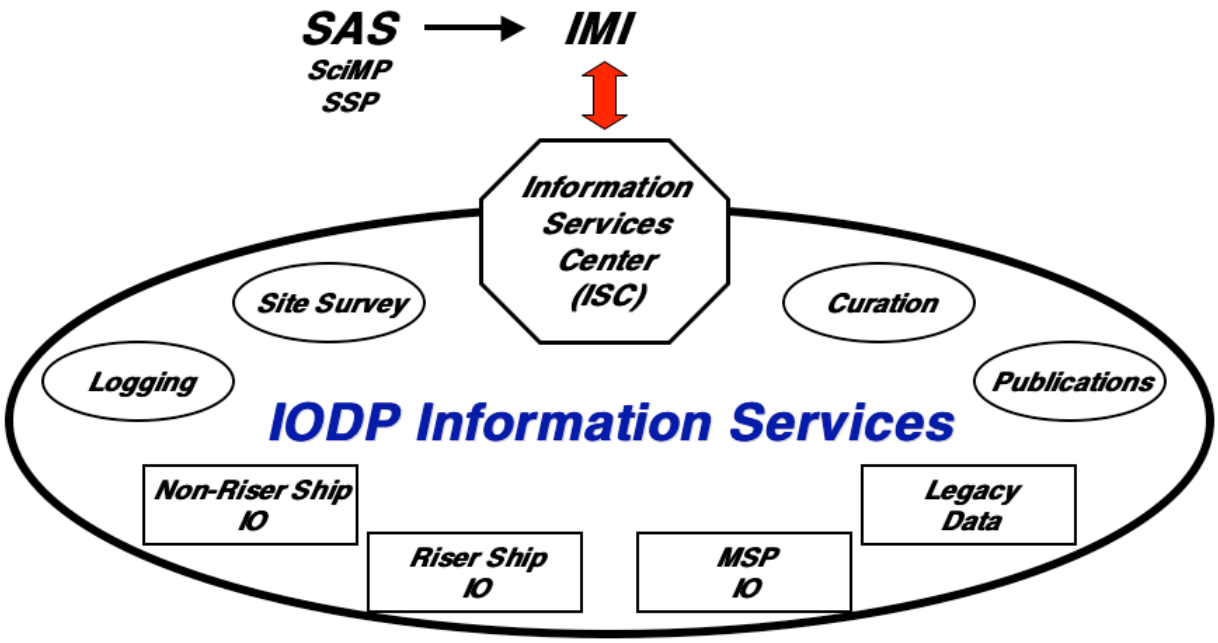

Figure 1. Proposed structure of IODP Information Services (IIS).

In a distributed environment, data resides on multiple computer systems in multiple formats at multiple locations. The challenge to the ISC will be to provide any data user a single point of entry into the myriad of IODP databases, text libraries, and catalogs (one stop shopping). In such a situation, the user relies on the **clearinghouse** to provide the access using simple point and click routines and a minimum of passwords. Thus, special computer programs (routines) need to be in place in order to access files, databases, catalogs, text libraries, etc. located on disparate computers around the world. This is a nontrivial task, to say the least. The Group felt that by identifying a **clearinghouse function** for the ISC specific tasks could be identified that would be the sole responsibility of the ISC, to which audits and performance measures could be made.

Since no contractual arrangements are envisioned between the ISC and IOs, the success of the ISC would rest, in part, on its ability to work in a cooperative sense with the IOs in order to successfully deliver information services to the scientific community. As such, the Group felt that a "dictatorial" (top down) management approach between the ISC and the IOs would not succeed. Rather, a **coordination function** for the ISC was envisioned as having a higher probability of success. To that end, the Group recommends the ISC take a proactive approach to establish data collection, storage, retrieval, and access standards with the complete involvement of the IOs. A SciMP subcommittee could be used to oversee this interaction.

2) IODP Information Services Center Responsibilities

The Group recommends that the ISC have the following specific responsibilities:

- *provide integrated access to all IODP data (e.g. shipboard and shore-based)*
- *develop & maintain:*

  - *the central program-wide web-based portal to stakeholders (scientists, educators, industry, policy-makers, public). Note: this portal should be dynamic & open to other international information systems & communities (e.g. physical oceanograph)*
  - *portal user interfaces that are scalable for different stakeholders*
- *following SAS advice, adopt & maintain standards to:*
  - *capture, storage, and distribution of data and metadata on each platform and of shore-based data. Required developments and implementations should be largely based on ISO, OGC, W3C standards and recommendations (for more information see http://www.fgdc.gov/standards/related_activities.html)*
  - *foster publication of data within IODP information services, e.g., using Digital Object Identifiers (DOI, http://www.doi.org)*
- *perform regular (360 degree) evaluations of the performance of the clearinghouse and the IOs in the delivery of IODP information services*
- *oversee the archiving of IODP legacy data (e.g, in partnership with recognized data centers)*
- *maintain and provide access to the program's publications database and integrate IODP information/data with IODP publications, e.g., using DOIs*
- *provide access to IODP curatorial information*
- *coordinate the development of data capture interfaces for specific platforms on an as-needed basis*
- *coordinate communications among the platform operator's IT/IS managers to share new ideas, resolve problems, and to adopt new information technologies.*
- *maintain links with other data groups (e.g. WDC, NGDC, ICDP, DEOS) and disseminate relevant information among IOs.*

Discussion:

The ISC should be the central location through which all publicly available IODP data and information are made available to IODP stakeholders. This is best accomplished through a portal that is both flexible and dynamic. The user interface should be scalable, that is, it should be able to accommodate both the novice and the experienced users, and most importantly, the user should always be able to find something related to their search. The portal will be based on levels of metadata, middleware, and user interface hardware and software. Implementation should be based on international standards (such as the ISO/TC 211 family of standards - http://www.isotc211.org/), which specify all necessary components for an effective geospatial data infrastructure, including "discovery", access, and exchange of IODP related data. Construction and maintenance of an IODP thesaurus, derived from metadata contents and related information inventories, will be one of the key elements to facilitate data and information access for the different stakeholders. By implementing such a design for its database management system IODP will be consistent with other oceanographic information systems, thus increasing the versatility and usefulness of IODP data for our understanding of the earth's systems and history.

The ISC should be tasked to follow the advice of the Science Advisory Structure (SAS) for the approval and adoption of metadata and data capture formats to be used on each of the operational platforms, as well as those formats used for upload of data sets into the IOs systems and

distribution via the ISC portal. This The ISC will maintain these standards and make sure that all data are accessible in the proper format. It is the adherence to agreed-to-standards that makes a distributed database management system work.

Regular performance evaluations should be carried out to determine how well the clearinghouse is meeting the needs of the IODP stakeholders and responding to their requests. The Group believes that this is an extremely important responsibility of the ISC. The ISC is a service organization and as such is responsible for providing information and data to the public in a form and manner that meets the needs of the public. Regular evaluations and reviews are essential to providing the best service possible.

IODP has spent time preparing for the beginning of drilling operations. The Group believes that now is the time to begin thinking about the end of drilling operations and providing for the legacy of IODP. There are many lessons to be learned regarding the preservation of legacy materials from the previous ocean drilling activities. Regular transfer of data to the appropriate archiving agencies during IODP should be the practice of the ISC in cooperation with the archiving agencies.

The data generated by IODP will include more than the data collected on the operational platforms. The data include "prime data" to be collected by IODP and then processed on shore, data published in the scientific literature, and publications that will be based on IODP data. The ISC should be charged with the responsibility to implement an information service that includes links to the publication information as well as access to the actual data. The Group recommends including Digital Object Identifiers to reference all IODP-related data publications. The DOI system would make data publications citable and thus provide credit to both IODP and the individual researcher, which would be mutually beneficial (The International DOI Foundation (IDF) and ISCU World Data Centers are currently piloting a project to investigate the premises for this procedure).

In addition to information describing the core material and the downhole environment, curatorial information should also be included in the information services system. Information regarding who has what samples, where those samples are from, and other similar information need to be included. The ISC could also be tasked to provide database support services to the IODP core repositories as would be appropriate.

Another specific ISC responsibility should be to coordinate database management activities of the ISC and the IOs. This coordination should include routine meetings between the IOs and the ISC to discuss system operation issues, new technologies, and new ideas. The ISC will also be responsible for interacting with the IOs to assure that all the necessary metadata are generated according to the agreed upon standards.

3) IODP Information Services Standard Practice

The Group recommends that IODP Information Services include the following standard practices:

- *The ISC should be regularly evaluated following IODP project management standards to ensure that it meets the data and information needs of the IODP stakeholders as defined by the SAS*
- *An annual review of the ISC by external IT/IS experts to ensure that IODP is utilizing the best technology possible (e.g. in terms of cost, applicability or efficiency)*
- *IOs should ensure that the standard (as defined by SAS) shipboard IODP data are captured electronically by the end of the moratorium period for each project*
- *IOs will work together with the ISC to provide consistent data collected on all platforms with particular attention given to common units, calibration information, and standardization of measurements (e. g. depth, age models, etc.)*
- *IOs are responsible for performing quality control and consistency checks on all data and metadata generated on their platform for each project*
- *The ISC will provide feedback to the IOs on the quality and consistency of the metadata supplied*

Discussion:

The ISC is, as its name implies, a service organization. Its primary function is to be the public image of IODP. It is where the public will go to receive information about the program, data from the program, and publications related to the program. These are very significant responsibilities. To maintain the high standards required to make IODP a premier science and world class research program, the ISC must successfully carry out its mission. In order to meet these responsibilities a minimum set of standard practices is recommended.

The Group's recommended standard practices involve both the ISC and the IOs. IODP will only be as successful as each of its individual components. The key is to measure or monitor the level of service to the public and the stakeholders. Regular evaluation of the service provided by the ISC should be performed by the IMI, with input from the SAS. This is essential to maintain high standards and expectations for the ISC. Additionally, a review of the ISC's technical capabilities by non-IODP technical experts is recommended. This review will address issues related to efficiency and technical operations of the Center. Both of these reviews will provide the ISC with the feedback it will need to assure that IODP is represented to its stakeholders in the best manner possible.

4) IODP Information Services Standards

Standards are essential to the success of the ISC clearinghouse. The Group recommends that:

- *Based on advice from the SAS, the ISC will adopt data standards for IODP consistent with international and emerging standards such as ISO and FGDC*

- *IOs provide the ISC with access to IODP data using consistent, standard metadata catalogues (e.g. in XML following adopted IODP standards)*

Discussion:

The SAS has a very important role in the design and operation of the ISC. The distributed system design should be built on accepted standards. This is valuable for two reasons; first, IODP is more likely to be interoperable with other large global oceanographic programs, and second and more importantly, legacy data are more likely to be compliant with search mechanisms and national archiving requirements. Adoption of standards thus fosters integration, widespread dissemination, and usage of IODP related data.

5) IODP Information Services Definition of Information

Information includes, but is not limited to:

- Shipboard and shore based collected data (ODP Janus data and microbiology, drilling parameters, downhole measurements, site-specific survey, paleontology, visual core description, XRF, CT data)
- Engineering data
- Citations that include IODP information
- Curation information
- Observatory data links
- Ship schedules
- Applications
- Project description information
- Policies
- Publications…..