

Data Management Coordination Group Meeting #2

Edinburgh, Scotland

June 27–30, 2005

Theme: Data Management in IODP

Host:

Colin Graham (BGS/ESO) ccg@bgs.ac.uk

Chair:

Bernard Miville (IODP-MI) bmiville@iodp-mi-sapporo.org

Main Participants:

Peter Blum	(TAMU /USIO)	blum@iodp.tamu.edu
Colin Graham	(BGS / ESO)	ccg@bgs.ac.uk
Shin'ichi Kuramoto	(CDEX)	s.kuramoto@jamstec.go.jp
Shigemi Matsuda	(CDEX)	matsudas@jamstec.go.jp
Rakesh Mithal	(TAMU /USIO)	mithal@iodp.tamu.edu
Emanuel Soeding	(IODP-MI)	esoeding@iodp-mi-sapporo.org
Michael Diepenbroek	(Pangaea/ ESO)	mdiepenbroek@pangaea.de

Core Repositories Representatives:

John Firth	(TAMU/USIO)	firth@iodp.tamu.edu
Kazushi Kuroki	(Kochi/CDEX)	kurokik@jamstec.go.jp
Walter Hale	(BCR/ESO)	whale@allgeo.uni-bremen.de

Guests/Observers (if available):

Hans Christian Larsen	(IODP-MI)	hclarsen@iodp-mi-sapporo.org
Frank Rack	(JOI)	frack@joiscience.org
Tom Janecek	(IODP-MI)	tjanecek@iodp.org
Ann Klaus	(TAMU/USIO)	annklaus@iodp.tamu.edu

Overview of discussion items

Data Management will be the main topic during this meeting. We will discuss both the IODP Data Management Working Document (to be distributed soon) and the future "Information Portal for IODP", (IPI, formerly known as ISC). We will also continue our development and discussion of the IODP Metadata profile and data model comparison exercise.

We would also like to start the discussion about data entry and visualization tools both onboard and offshore. Representatives from the Core Repositories are also invited, as we will talk about core redistribution from the perspective of data management and the capturing and management of post cruise data.

Main Agenda Items

1. IODP Data Management Plan

The IODP data management plan will be delivered at the beginning of June. The plan will include discussion about all type of IODP data, from drilling to reports including minimum IODP data measurements. The focus will also be on the integration of all IODP data into a simple one access system (discovery portal and future viewing options) for scientists (Information Portal for IODP, IPI).

We would also like to discuss the road map for the IPI, its possible content and how we see the implementation process.

2. IODP Metadata Profile

We already decided on a metadata standard and discussed the structure and content. A preliminary metadata profile was created. A future schema will also be developed pending the release of the ISO 19139 Implementation standard. At this meeting we would like to discuss any final adjustments

needed to the proposed IODP metadata profile. We also need to discuss how the IOs will provide the required metadata to the future metadata catalog (part of the IPI).

We will also start the discussion on discovery level metadata for all IODP products (images, reports, publications, post cruise data, etc.) and metadata for quality control purposes.

3. Datamodels

A preliminary study of the difference between the JANUS, J-CORES and DIS data models was initiated. The main purpose of this exercise was to understand the difference and see if it is possible to map one data model to the other. We will discuss the results of this comparison and the consequences of the differences.

4. IODP drilling data viewing and manipulation tools

We would like to start the discussion about data entry, visualization and manipulation tools. Basically we need to find out what kind of tools are currently in place both on board and offshore and how they are used and where they are located in the process of gathering drilling data. We also need to know any tools that currently exist and might not have been tested and what other tools are being developed or wished for by the IOs.

We would like to have a list already in place before the meeting. IODP-MI provided an Excel sheet to all IOs, on which information is requested. The result will be presented at the meeting.

This topic will require the participation of all 3 IOs via a presentation or demonstration of their current and planned visualization and manipulation tools. We are also interested in finding out whether any of these tools can be used or modified to accept a common agreed input data standard (XML format?). If we all agree on an input standard for the tools, any tools developed by any IOs should work on any platform and be interchangeable. We do plan to start the discussion in Edinburgh, which will be continued and refined on subsequent meetings.

We would like to start the discussion on how to solve the Lithology problem at this meeting, with discussions on how to involve scientists and the community in general.

This is another area where collaboration between the IOs will become important.

5. Core repositories

The geographically based redistribution of cores in FY06-FY07 requires action to be taken. We would like to discuss any potential problems this redistribution with regards to data management (measurements, sample requests, etc.) could cause.

We also need to discuss how to better capture and manage post cruise data.

We would like to see a short presentation from the 3 curators/representatives about their work and their perspective on possible problems with the data management of geographically based distribution of cores. We would also like to see their suggestions on how to capture and manage post cruise data. A break out session chair by Hans Christian Larsen will take place in the afternoon of the first day to further discuss all issues related to Core Repositories.

Other topics

1. STP request on QA/QC

STP has issued a request to the IOs to comment on QA/QC protocols and their implementation (SciMP Action item 0502-11).

Action item 0502-11: SciMP recognizes the need for QA/QC protocols to be implemented for all scientific measurements to be made by IODP. In consultation with the IOs, each SciMP WG explicitly prepares a draft list of the existing or planned apparatus, their QA/QC and calibration procedures and reference materials used in their specialty areas. Each working group will evaluate the QA/QC protocols and make recommendations for their implementation that will:

- 1) be uniform across the different platforms;
- 2) be routinely used;
- 3) be sufficiently flexible to meet specific expedition scientific goals while maintaining the high quality of the data produced;
- 4) Allow easy comparison of similar data recorded by different platforms.

Each working group will report their findings at the next SciMP meeting.

Action to be taken by: SciMP (Neal, Korja, Suzuki, Villinger, Kenji) and IOs (Blum, Roehl, Kuramoto).

During this topic each IO should briefly explain QA/QC implementation. A routine procedure should be outlined, to prepare a sufficient response to the respective STP working groups.

2. Error ranges

If and how are error ranges recorded with the data? How can this be implemented into the database system. This item overlaps with the QA/QC issue. One Question to address is, where these errors should be recorded, with the data in the measurement database, or with the QA/QC data in the instrument database?

Meeting logistics

There is transport (bus) arranged to take attendees of EPSP, DMG and OPCOM from the hotel to BGS and return in the evening. There are probably about 40 people to transport on the Monday to Wednesday - less on Thursday. The bus times are arranged so that meetings can start in BGS at 9.00am and finish at 5.30pm.

During break, tea, coffee etc will be available plus snacks, all provided free by BGS.

Buffet lunch (free) will be provided in a canteen on the university campus (about 5 minutes walk from BGS).

There will be a white board, projector and telephone, plus access to power (and a few plug adapters). One of 2 Mac's video cable adapter will be available. You will not be able to connect your laptops to the BGS network because of the firewall- but BGS will provide a number of client PCs so that attendees can access the Internet and printers.

Meeting agenda schedule

The meetings will take the full first 3 days to discuss the main agenda items and on the fourth day we will try to start the discussion related to QA/QC and Lithology issues. The proposed agenda schedule follows:

Monday, June 27 2005

Time	Topic	Presenter
9h00	Welcome, Introduction, agenda, logistics	BM
9h30	Summary of action items	BM
10h00	Summary about IODP Data Management	BM
10h30	Break	
11h00	Gulf Core Repository*	JF
11h15	Bremen Core Repository*	WH
11h30	Kochi Core Repository*	KF
11h45	Core Repositories Discussion	BM, HCL, DMCG
12h30	Lunch	
13h30	Core Repositories Break Out Session (Separate room)	HCL, ES, JF, WH, KF, Tom Janecek**, Ann Klaus
13h30	Data Models Comparison - Update - Issues	BM, DMCG
14h30	Discovery level metadata - Update - Issues	BM, DMCG
15h30	Break	
16h00	- Discovery level metadata for other data type (initial discussion) <ul style="list-style-type: none"> o Images o Reports, Publication o Post Cruise research data o Others - Metadata for QA/QC, how to proceed	BM, DMCG
17h30	End	

*: Short Presentations about:

- Data Management System used
- QA/QC
- Core redistribution potential problems
- Post expedition research data tracking and archiving

**: If available

Tuesday, June 28 2005

Time	Topic	Presenter
9h00	Introduction Data tools	BM, ES
9h45	Data tools (J-CORES, SIO7)*	CDEX
10h30	Break	
11h00	Data tools (DIS/PANGAEA)*	ESO
11h45	Data tools (JANUS)*	USIO
12h30	Lunch	
13h30	Discussion <ul style="list-style-type: none">- Tools needed (data input and viewing)- How to collaborate- Testing CDEX tools on JR- Tools data input standard, connection to databases, how and where to start- Tools for QA/QC	BM, ES, DMCG
15h30	Break	
16h00	<ul style="list-style-type: none">- Discussion continued- Conclusion, future action items	BM, ES, DMCG
17h30	End	

* Presentations about:

- What they already have
- What they are developing
- What they would like to see
- How they see collaboration between the IOs

Wednesday, June 29 2005

Time	Topic	Presenter
9h00	IODP Data Management Working Document <ul style="list-style-type: none">- Goals- Process- Assessment- Proposed solution	BM
10h30	Break	
11h00	IODP Minimum Measurement Data <ul style="list-style-type: none">- Issues- Impact on discovery level metadata- Action Items	BM, DMCG
12h30	Lunch	
13h30	Information Portal for IODP (IPI) <ul style="list-style-type: none">- Content- Discovery level metadata harvesting for IPI- Implementation	BM, DMCG
15h30	Break	
16h00	<ul style="list-style-type: none">- Inter Database Data migration- Other topic of discussion	BM, DMCG
17h30	End	

Thursday, June 30 2005

Time	Topic	Presenter
9h00	<ul style="list-style-type: none">- Summary of QA/QC issues- Document for SciMP?- Decide on what to do	ES, DMCG
10h30	Break	
11h00	<ul style="list-style-type: none">- What do we want from STP about Lithology- Decide on what to do	ES, DMCG
12h30	Lunch	
13h30	Open Discussion as needed	DMCG
15h30	Break	
16h00	<ul style="list-style-type: none">- Future action items- Summary and wrap-up of meeting	BM, ES, HCL, DMCG
17h30	End	

Draft Summary Report and Action Items from Data Management Coordination Group Meeting #2 Edinburgh, Scotland

June 27–30, 2005

Theme: Data Management in IODP

Host:

Colin Graham (BGS/ESO) ccg@bgs.ac.uk

Chair:

Bernard Miville (IODP-MI) bmiville@iodp-mi-sapporo.org

Main Participants:

Peter Blum (TAMU /USIO) blum@iodp.tamu.edu
Colin Graham (BGS / ESO) ccg@bgs.ac.uk
Shin'ichi Kuramoto (CDEX) s.kuramoto@jamstec.go.jp
Shigemi Matsuda (CDEX) matsudas@jamstec.go.jp
Rakesh Mithal (TAMU /USIO) mithal@iodp.tamu.edu
Emanuel Soeding (IODP-MI) esoeding@iodp-mi-sapporo.org
Michael Diepenbroek (Pangaea/ ESO) mdiepenbroek@pangaea.de

Core Repositories Representatives:

John Firth (TAMU/USIO) firth@iodp.tamu.edu
Kazushi Kuroki (Kochi/CDEX) kurokik@jamstec.go.jp
Walter Hale (BCR/ESO) whale@allgeo.uni-bremen.de

Guests/Observers

Hans Christian Larsen (IODP-MI) hclarsen@iodp-mi-sapporo.org
Frank Rack (JOI) frack@joiscience.org
Tom Janecek (IODP-MI) tjanecek@iodp.org
Ann Klaus (TAMU/USIO) annklaus@iodp.tamu.edu

Executive Summary

The overall IODP Data Management was the main topic of the Edinburgh meeting. The main topics discussed were:

- IODP Data Management
- IODP data curation
- IODP Data model and Discovery level metadata profile
- IODP Data tools, and data format
- Information Portal for IODP (IPI)
- IODP Data definition (Minimum measurements)
- Data tools test on JR
- Lithology
- QA/QC

The "IODP Data Management Working Document" was discussed and in principle everybody agreed on the process, not everybody could agree about the details. In particular there was discussion about the role of PANGAEA in the overall IODP picture. That discussion continued after the meeting within a series of e-mail exchanges.

Curation was discussed in break out sessions from the DMCG meeting. Everybody agreed on the need for an overall IODP Curation System that all IO's would use. The draft requirements are currently being written by TAMU (JF).

The IODP metadata profile will move into another discussion period via e-mail to ensure that the content fits well with the actual need of searching and finding IODP data. Once the discussion is completed, the metadata schema will be finalised and IO's will be requested to provide metadata in the specified format for the future Information portal for IODP (IPI).

The content and possible format for the future IPI was presented by IODP-MI and all agreed on a modified scenario, but there are still discussions on the role each IO's database will play.

Existing and future application tools were discussed and how the three IO's could collaborate in improving and making them interoperable, portable and open source. A list of possible tools were agreed and it was decided to start with one existing tool (VCD) and one future tool (IODP Curation System). The development of the tools involves also deciding on the method to input and export data. One of the main issues will be to develop common exchange data format that everybody will use.

Some of the tools will be tested during two weeks of transit on the JR in September 2005 In which all three IO data managers and specialists will participate.

IODP minimum measurements were discussed and it was decided to propose a process to STP. The process involves creation of a master list of measurements and categorizing them into different level of requirements depending on the type of expedition. A memorandum was written and passed on to STP for comments at their July 2005 meeting.

A similar process is envisioned for the lithostratigraphy, where the DMCG will propose a draft to STP. Part of the test on the JR will probably be used for that discussion.

Discussion of the QA/QC issue was initiated at the meeting. Prior to the meeting, IODP-MI had requested lists of instruments, tools and software used during an expedition and within core repositories. That list will be re-organized and for each instrument several characteristics will be tabled including QA/QC procedure.

All participants are thanked for contributing to a very productive meeting. In the following sections more details about each topic and actions items are presented.

1. IODP Data Management

Just a few weeks before the meeting, the DMCG members received with the "IODP Data Management working document" (for now on IDMWD) where different IODP data goals and the "Information Portal for IODP" (for now on IPI) were presented. The main purpose of the meeting was to discuss some of these goals and the IPI and to see how we can start integrating all the IODP data access, collection and tools. We had several productive discussions about data management and several other related topics, but we are still failing short of any consensus. Currently the IODP data flow is like in figure 1:

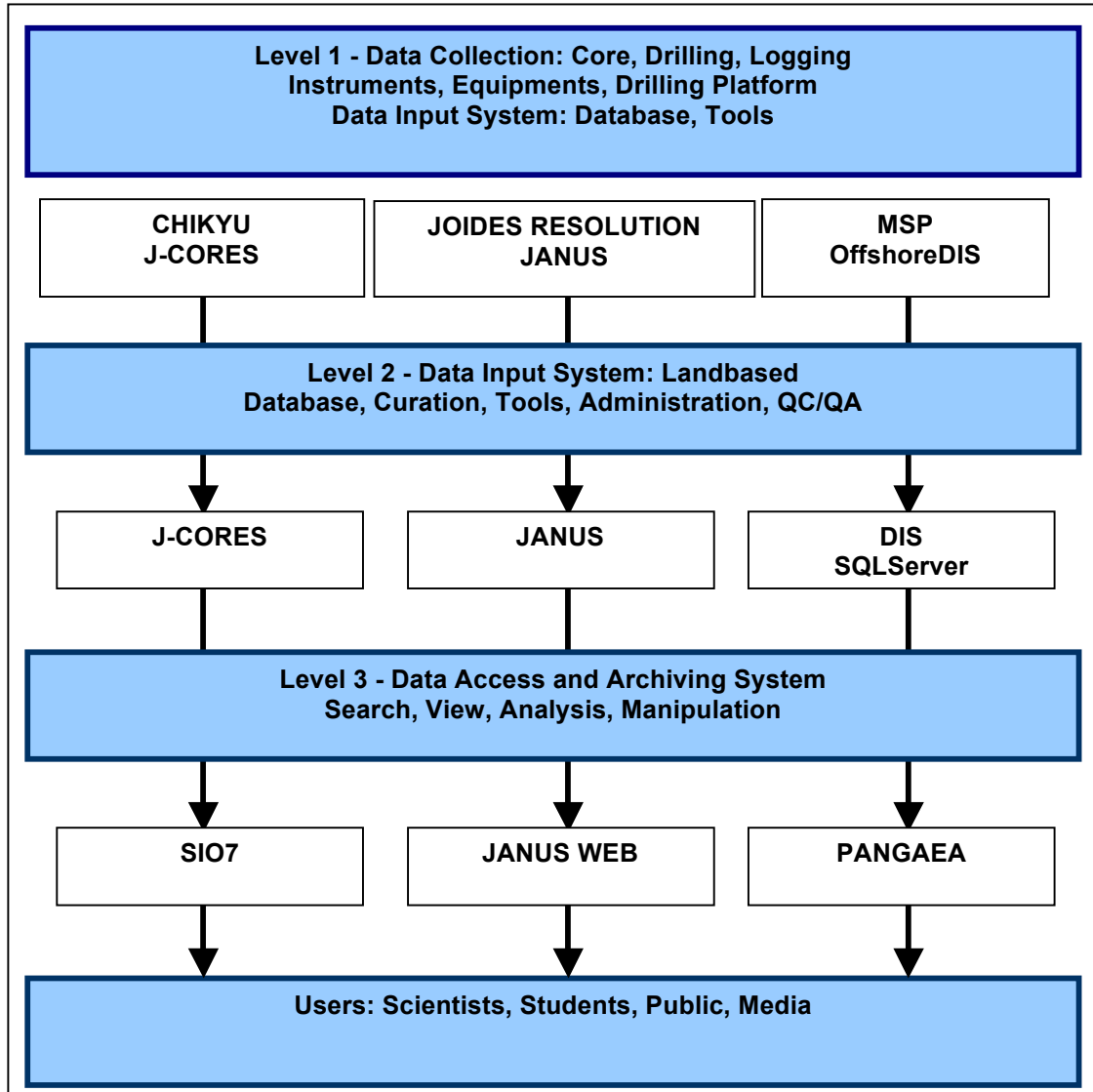


Figure 1: Current Data Management Situation

It is simple but not very integrated and cumbersome for the user to find any data. After discussion at the Edinburgh meeting we adjusted the proposed data flow brought up in the "IODP Data Management working document" for integration to a new IODP data flow as in figure 2. It includes the integration using common metadata and data format with the possibility of using some common data entry and manipulation tool. Collaboration is the key issue, where we would envision all IO's working together in the development of metadata, data format and common tools.

There are still some discussions about the role of each individual database but hopefully we can come up with a consensus in the near future.

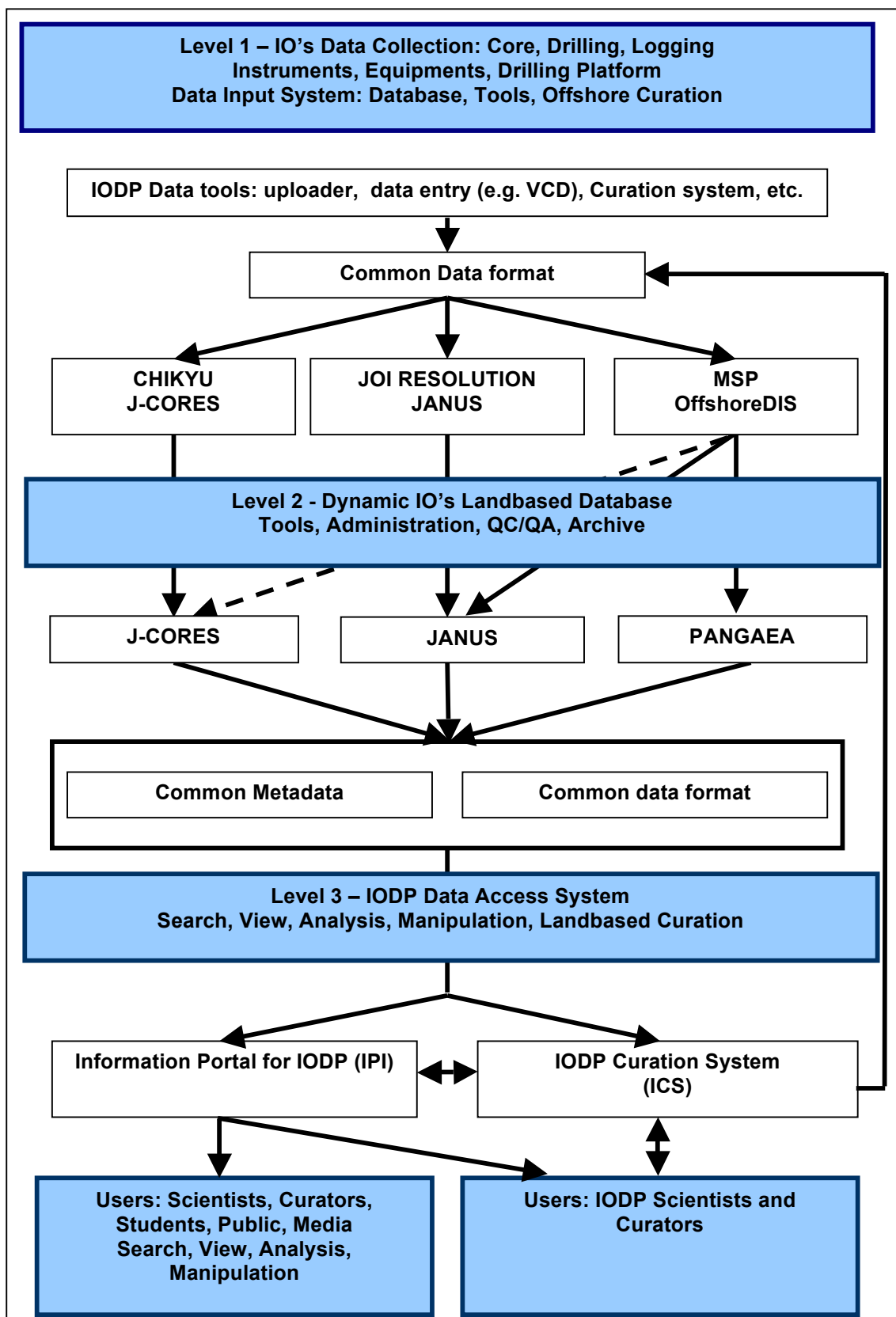


Figure 2: Suggested Future IODP Data flow

In the figure there is a needed link between the Offshore DIS and JANUS. This is hopefully temporary needed until the “IODP Curation System” is created and functional. Until that time, MSP data will need to be migrated to JANUS for sampling purposes.

Once we have a consensus, IODP-MI will write up a revised version of the IDMWD to reflect the new agreements and goals and will distribute it for comments.

Action Item

- IO's Manager to come up with a consensus about IODP Data Management, Hans Christian is leading that effort
- Once we have a consensus, IODP-MI will revised the IDMWD
- Make sure MSP data continues to be automatically transferred to JANUS until a functional IODP Curation system is in place.
- Fully understand PANGAEA capability in comparison to JANUS and J-CORES for possible future integration in the Information Portal for IODP

2. Core repositories

We had presentations from the 3 main curators in IODP and John Firth gave an extensive presentation that started a lot of useful discussion.

There were several break out sessions about repository and curatorial policy issues with the following participants:

- Hans Christian Larsen
- Ann Klaus
- John Firth
- Kazushi Kuroki
- Walter Hale

The result of the discussion is in appendix A at the end of this report.

There was also a lot of discussion on the need for a unified and single curation tool for IODP. The curators went again in a break out session to define what they feel is needed in a possible curation tool.

There was a general consensus that there is a clear need for a common tool for the whole curation process. That includes:

- Sample request
- Sampling on ship and at the core repositories
- Post expedition sample data

That could be achieved by creating a single IODP Curation System (ICS). That will involve a single database for sample id and science data. This system would be the only entry place where sample data could be searched and entered. This system would allow someone at a core repository to search and be aware of all the samples no matter where the sampling has been done or where the physical sample data is being stored.

It would also involve synchronization between the land based sample system and the ship based where most of the expedition sampling may take place.

The initial requirements for an ICS were written in a break out session at the DMCG meeting between the 3 curators. A more formal document will be written by John Firth and technical information will be added by IODP-MI.

Action Item
<ul style="list-style-type: none">- John Firth will write an initial document including the specific requirements for an ICS (4-6 weeks)- This document will be distributed to DMCG and curators for comments- IODP-MI will add technical information for the system- The document will be distributed again for comments- Final version created by October and ready for discussion on implementation

The basic idea behind the system is a single entry point for managing and searching. Sample ID requires knowing Core log data. Currently only the JANUS database is holding that information. If we are going to create a new IODP curation system, it will need to be able of gathering Core log data from multiple databases (JANUS, J-CORES, and PANGAEA). There will be a need to access the capability of each database to provide the required info. If they can not, so alternative method of gathering the needed data will need to be proposed.

One possibility is to make the ICS part of the IPI. Since the IPI will already be harvesting discovery level metadata from the 3 main IODP databases, it would just make sense to harvest core log data for curation at the same time. Since Internet connection at sea can not be always guarantee, the ship board ICS would need to be synchronized before leaving for an expedition and whatever was collected during the expedition would need to be fed back into the main ICS.

3. Data Models

The data model exercise was briefly discussed and it was agreed that it was useful in recognizing the differences between the 3 IODP core data databases and that it could be use in the future to map data from one system to the other. No further action items are required at the moment.

4. IODP Metadata Profile

The metadata profile that has been developed over the past few months, need to be finalized. There were concerns about the level of granularity we originally decided to describe IODP data. Including metadata at the discovery level for sample data was argued by Michael as being too small entity and not reasonable because we would end up with too many sample metadata very quickly making the system inefficient for searching. It was also argued that typically a researcher would not look for sample data but would generally want to find out which expedition/core/hole measured a certain type of data.

On the other side it was argued that sample is the main data of IODP and metadata needs to be provided to the future IPI. So no agreement was reached at this level and the discussion will continue via the DMCG e-mail list.

Basically we need to make sure that the Metadata is closely linked to what is actually going to be searched on. Several search domains were discussed:

- **Discovery Level**
 - Scientists Request
 - who, where, when, what, how, why
 - Cross-Metadata Queries
 - What data is available?
 - What tools are there?
- **Curatorial Data (Sample data)**
 - Scientist
 - Curator
- **Expedition Data**
 - Scientist
 - Curator

Based on these scenarios, we agreed that not matter what metadata for sample will be needed, but it might be different than discovery level metadata.

Action Item

- Need to (define subgroup to) come up with Data Request Scenarios
- Continue the discussion based on search scenarios on the content and granularity of IODP discovery level metadata profile via the DMCG e-mail list. IODP-MI will send some suggestion.
- Discussed the need for different sample data metadata for the future Curation System
- Once we have a consensus on the content and granularity for the metadata, the schema will be created by IODP-MI
- After the schema is created, the IO's will be able to start working in providing metadata to the IPI

5. IODP data entry, viewing and manipulation tools

The 3 IO's presented information about their data tools.

CDEX did an extended presentation that generated a lot of interest especially with the Visual Core Description (VDC) tool. CDEX mentioned that their software will be open source.

ESO did a demo of their implementation of DIS with some impressive set of tools and a very flexible database management (SQLServer) that can be adjusted quickly to any drilling platform and instruments. However they do not plan to develop any new data tools.

USIO mainly talked about their LIMS software which also generated interest in being implemented in all IO's.

Michael did a presentation about PANGAEA, showing how interconnected it is with all other Geosciences portal. We also saw the infrastructure of PANGAEA and it's already in place capability to provide metadata using ISO 19115 standard.

The main purpose of this topic was to identify tools where collaboration in the development between the IO's could be possible. This collaboration is needed to avoid duplication of effort and ensure that scientists and users all use the same tool for doing the same work. An initial list of the most important data tools where collaboration would be desirable was created at the meeting (see Table 1).

Table 1: List of data tools where collaboration would be desirable

Category	Tool	Type	Comments
Visual Core Description	VCD (J-CORES)	Data Capture to DB, Viewing, Manipulation	In development Hard coded for J-CORES Terminology Test
	AppleCore (TAMU)	Data Capture to text file and binary	In place, Commercial software
	DIS	Data Capture	In place
Sampling	Sampling (TAMU)	Data Capture	In place
	DIS (ESO)	Date Capture	Hard coded for own DB
	Sample (J-CORES)	Date Capture	Need to be sync with other DB
Core Log Viewing	CLV (J-CORES)	Data viewing	In development Hard coded for J-CORES
Depth (age) Correlation	Splicer/Sagan (TAMU)	Data Manipulation Data Capture	In place Ported to JAVA NFS proposal
	D-Tunes (J-CORES)	Data Manipulation Data Capture	In development
Instrument Management	LIMS (TAMU)	Data Capture and access	Future development
Paleontology	Paleo (TAMU)	Data Capture	In place but not used much on ship, but used onshore Hard coded for JANUS
	Stratigraphy (J-CORES)	Data Capture, viewing	Hard coded to J-CORES
Sample request	Various	Info capture	Coordinate collection of info

We also discussed and agreed what should the basic requirements for the future IODP data tools should be:

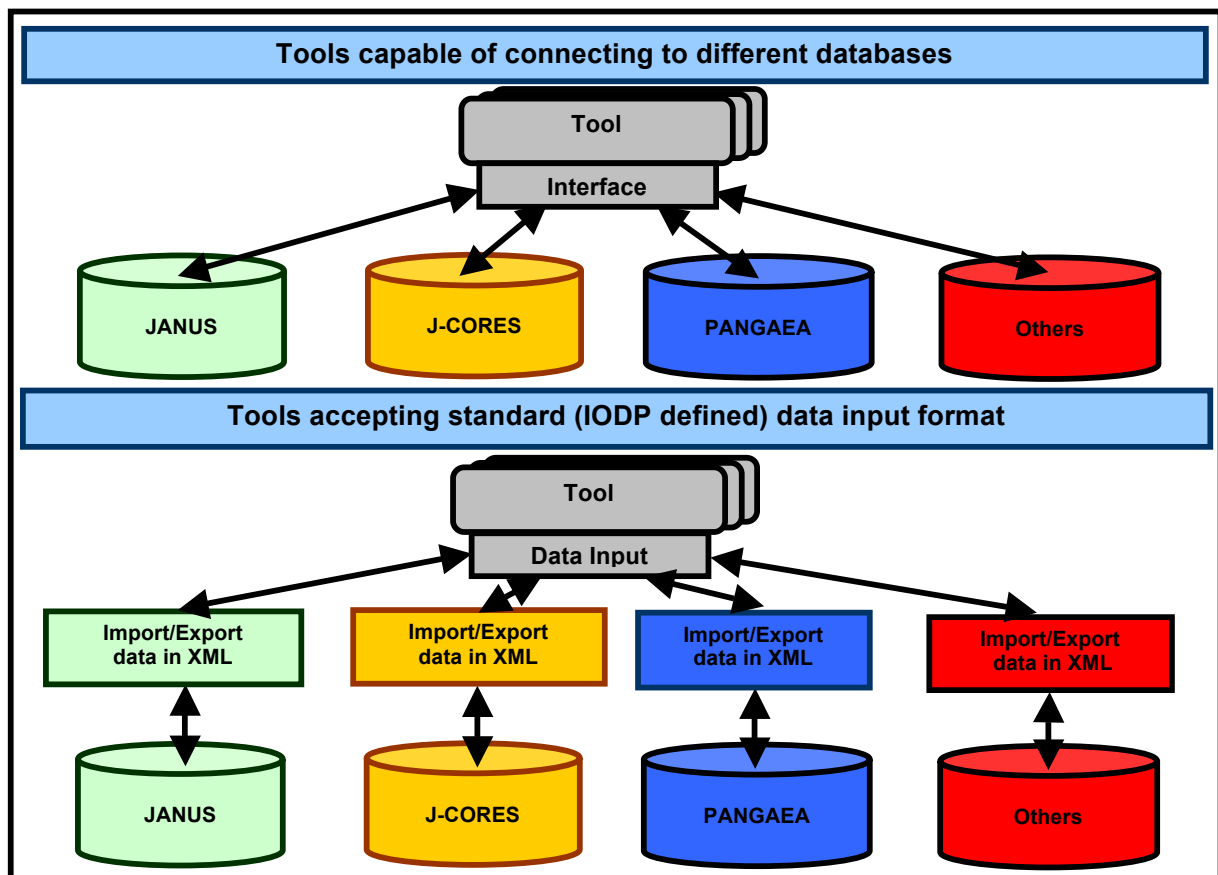
- Interoperable
 - Works on data coming from any database
 - Works on all computer platform and operating system

- Portable
 - Works anywhere no matter if it is connected to a database or not, works on local data
- Make software Open Source
 - Initially open to internal IODP community
 - Once something developed, open to external IODP community
- However everything depends on the availability of resources

The method for these tools to input and export data has not been fully discussed at the meeting. More information on existing system needs to be looked at in order to properly access the resources needed to developed or modified current data tools.

Most data tools currently connect only to one database, no formatted data is actually being exchanged only direct data extraction and writing to a database. This makes the tools very restrictive and can not be used with any other data and the only way to use the tool is to migrate data into the connected database. It is possible to configure the tools to access multiple databases, but again this assumes 100% connectivity to a network.

If we want fully interoperable and portable tools, they will need to be capable of receiving formatted data that could come from anywhere regardless of their original storage system. This has the distinct advantage that the data could reside on your PC hard disk or from a database that can export in the data format required by the tools. It also goes the reverse way where data export from the tool would need to be formatted in a common format so that other tools or database could import the content.



There are currently several tools that are under development and it was decided at the meeting to start with at least one tool and see how the process goes. From CDEX, we will start with the VCD. The whole process will first involve looking at the initial specification as created by CDEX and access if it meets all IO's requirement in terms of functionality. The second step will be to actually test the VCD on the JR. Once that is done we will need to write a modified set of requirements that will accommodate all IO's need. CDEX is very open in participating in this process. The only issue is they would prefer to wait until the software has actually been delivered before making any modification because of the

contract they have with the developing company (September 2005). CDEX mentioned that they will be very busy for the next year and collaboration in terms of modifying the tools will be difficult. However once the software is delivered to them, it will be open source for all IO's to look at. This tool has obviously great potential of being used by all 3 IO's.

We also agree that a common curation tool is needed to manage everything from request, sample ID and science data entry. More details about the creation of a curation system are available in section 2.

This involves several action items as listed below.

Action Items

- Obtain original specifications for VCD (J-CORES)
- Review VCD Specification and modify to accommodate all IO's need
- Create specification for IODP Curation Tool - ICT (JF)
- Test functionality of CDEX tools on JR (August 29 to September 16)
- Obtain more information about 2 possible data tool connectivity solutions
 - Via data exchange format
 - XML formatted data from Oil industry
 - Marine XML
 - Via direct database connectivity
- Based on review, agree on best scenario for IODP (XML or database connectivity)
- Develop scenario for collaboration: How, who does what, workshop, meeting, terminology, control vocabulary definition for some tools, etc.
- Will have to be based on resources available

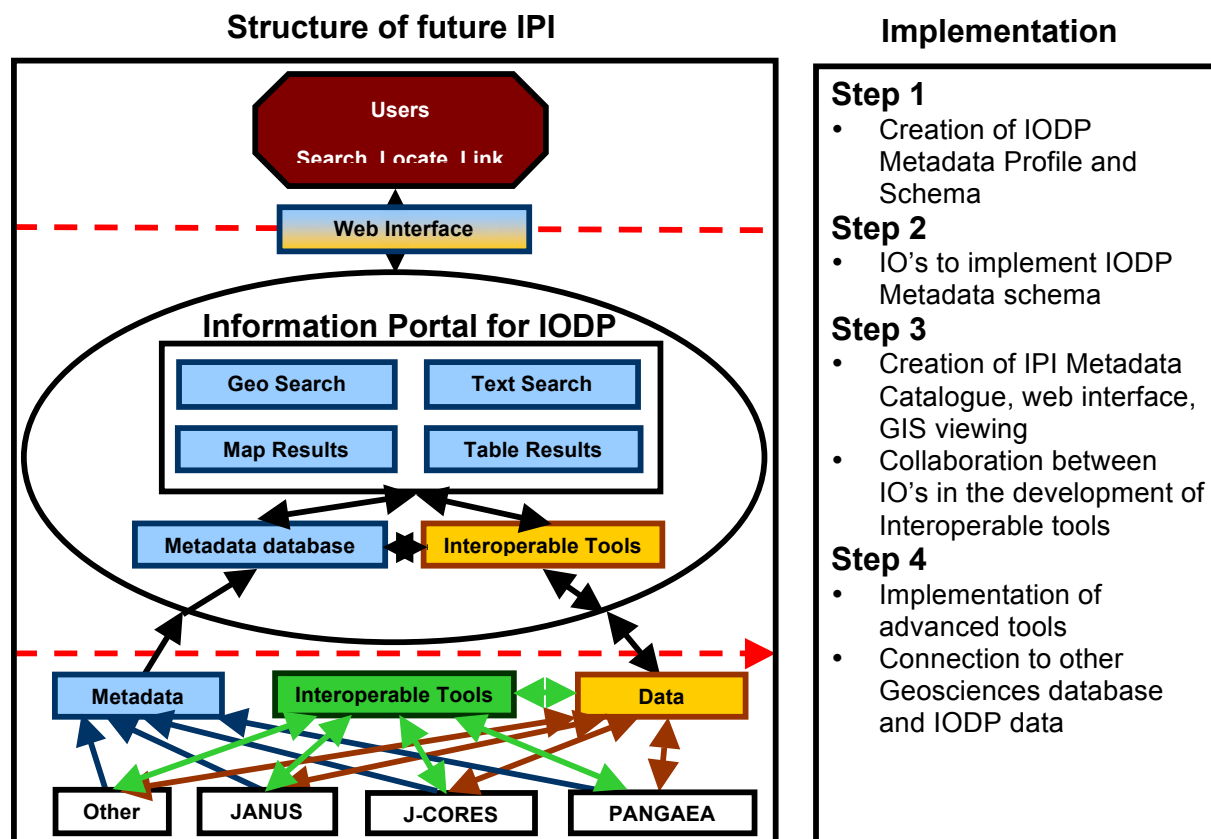
6. Information Portal for IODP (IPI)

The basic idea of an “Information Portal for IODP” was presented at the meeting. There was a general agreement to move forward with the concept but the details need to be worked out. The main reason for the creation of a portal is to make the access to IODP data easy for the users, especially now that we have 3 implementation organizations using 3 different data management system. We want the user to be capable of going to a one stop web portal where all IODP data can be search and found no matter where it is located.

In order to integrate all IODP data without imposing one data management system from the data collection offshore to the data access, all data needs to be integrated at some level. For discovery level type of portal, the use of common metadata for describing IODP data is the most efficient method to use. That implies that the IO's will need to provide the metadata to the future IPI (figure 2).

Based on the discussion about collaboration in the development of data tools, the figure includes the use of common tools from the data collection level to the curation at core repositories. None of these tools currently exists but with collaboration between all IO's, we should be able to move forward and create IODP wide tools.

The initial IPI will be built on just the current ODP/IODP data located in JANUS, J-CORES and PANGAEA.



Once the system has been established, it will be possible to expand to other related drilling data. The main purpose of the system is to avoid having users search multiple databases to find all IODP and legacy data. This is one level of integration that was agreed between all participants.

The system is completely dependant on Metadata being provided by IO's and other Geosciences databases in the future. The expanded version will eventually allow user to find data using web enabled interoperable tools.

Here again the first step is to finalize the IODP metadata profile and provide a schema to the IO's so the metadata can be created. The metadata would need to follow the IODP schema and be formatted in XML. Several methods for the creation of the metadata are possible:

- Write the metadata dynamically to files on IO's servers
- Write the metadata dynamically into blob in the IO's relational database

The preferred method for the IPI to access the metadata is to harvest it on a regular schedule. Again several methods are possible. IODP-MI suggests using existing software and protocol and that the most suitable method is probably the Open Archive Initiative Protocol for Metadata Harvesting:

<http://www.openarchives.org/>

Content of future Information Portal for IODP (IPI)

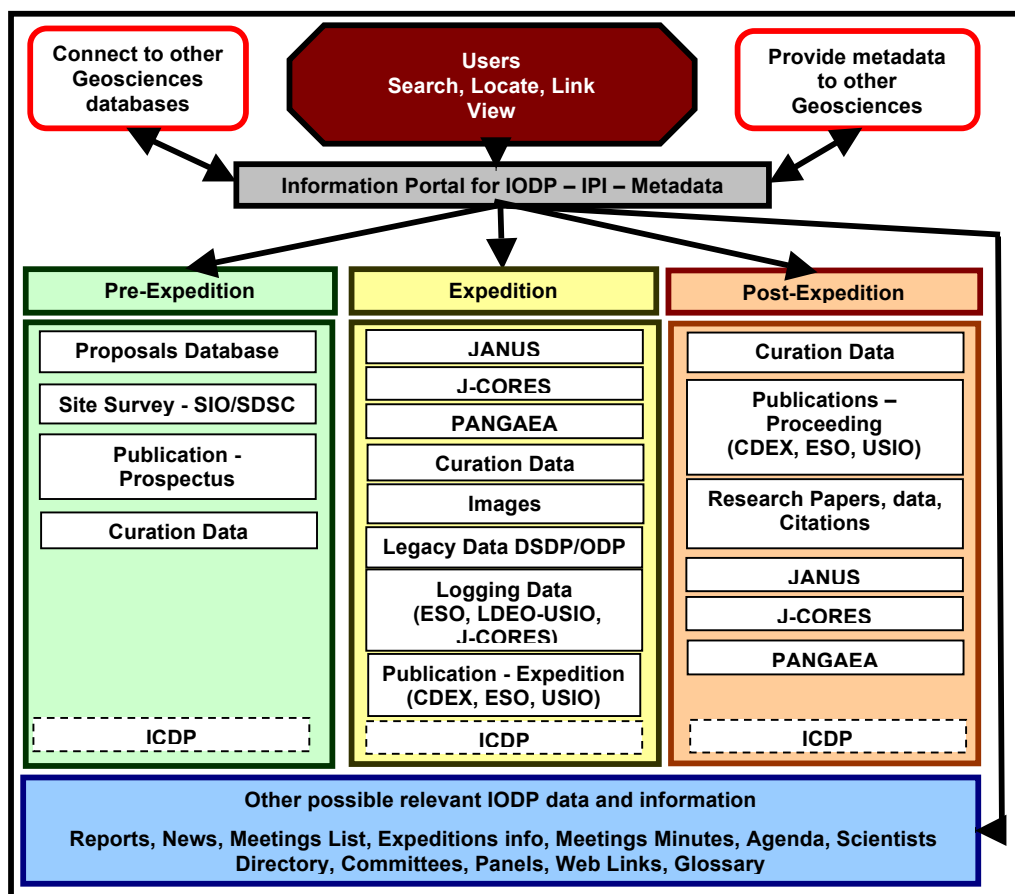


Figure 3: Expected content of future IPI

Creation of a task force will probably be needed for the development and monitoring of the IPI. Define a subgroup to define possible searches in metadata/data. All agreed that we need metadata on sample level but no necessarily for the IPI, but definitely for the Curation System.

Action Items

- Need to finalize IODP Metadata Profile first and create schema
- IODP-MI will suggest process for developing IPI (Detail IPI requirement document by IODP-MI)
- IO's will implement metadata schema

7. IODP data definition

An initial document about minimum measurement was produced in 2003 and there was no follow up. The DMCG looked at the document and decided to initiate an action item. It was decided to propose a process on how to categorize and define IODP data not only minimum measurement. An initial memorandum was written and with the agreement of all the IO's it will be presented at the upcoming STP meeting.

The memorandum is in the appendix B section of this document. Basically we are suggesting that a master list of all measurement types that are possible be created with clear definition for each type. Once that is created, we suggest that each measurement be categorized in one of the following group:

- Safety
- Minimum
- Standard
- Other

More specific information about these 4 groups is available in the memorandum.

We are hoping to have STP approval for the suggested process in order to have this issue moved forward.

Minimum Measurements can be categorized as:

- Level 1: Measurements for Safety and Operational Issues (POC)
- Level 2: Minimum Scientific Measurements
 - o 2a: ephemeral Measurements (needs to be done on the platform)
 - o 2b: non-ephemeral Measurements (can also be done in the shore lab)
- Level 3: Standard Mission Specific Measurements
- Level 4: Other Measurements

Action Items

- Define what is Minimum Measurement
- Ask Pubs people to define the term "Prime Data"
- Need to make a matrix of Tools vs. purposes / levels / platforms, to define where the tools are relative to the levels mentioned
- Let STP know that and how we are working on this (Memorandum attached)
- Let STP know that the tool list they requested can not be delivered easily immediately, but we are working on coordinating it

8. Test of data tools on JR

Data tools, from entry, visualization and manipulation will be tested on the JR on a transit cruise from the approximate dates:

Departure: Panama

Date: August 28, 2005

Arrival: Astoria

Date: September 14, 2005

The main J-CORES data tools that will be tested are:

- Sample Curation Operation
- Composite Log Viewer (CLV)
- Stratigraphy
- VCD
- Uploader

There will also be demonstrations of some tools from TAMU (e.g. Splicer, Sagan) and ESO (e.g. Offshore DIS).

The following people will participate in the cruise:

- Bernard Miville (IODP-MI)
- Manu Soeding (IODP-MI)
- Matsuda-san (CDEX)
- Takahashi-san (CDEX)
- Rakesh Mithal (USIO)
- Peter Blum (USIO)
- Jay Miller (USIO, staff scientist)
- Paul Foster (USIO, Application Developer)
- John Firth (USIO, Curator)
- Carlos Alvarez (USIOI)
- Colin Graham (ESO, depending on Tahiti cruise)
- Dave McInroy (ESO)
- Walter Hale (ESO, to be confirmed)

Action Item

Prepare an Excel based check list for functionality and comments. IODP-MI will create a first draft.

Action Item

USIO is going to organized all the logistics for the ships itself, accommodation, food, etc. USIO will also bring some physical core and samples from expedition 206.

Action Item

Matsuda-san will organize a daily testing schedule and bring some data files from their instruments to test the uploader. CDEX participants will probably be at TAMU one week before the cruise for last minutes preparation.

9. STP request on QA/QC

The QA/QC topic was discussed many times in other related topics during the meetings.

We agreed to come up with a joint IO's answer to the STP Action Item 0502-11: QA/QC protocols. The joint IO's input to this STP request will consist of parts of the discussions from the other topics in this meeting. This joint effort will be initiated during the J-CORES test cruise this coming September as most of the DMCG members will be present.

Action Item

- To discuss topic during J-CORES test cruise

10. IODP Lithology and Visual Core Description (VCD)

A structured handling of VCD data has been under discussion for a long time. This is an important issue, where standards should be set throughout the IODP program, to maintain comparability of data. It is however not clear how to approach this problem, and how to define standards, that will be acceptable to the majority of the scientists. *DMCG will therefore come up with a request to STP, to devise a uniform IODP lithological classification scheme. In order to make core data searchable and comparable within a database.*

Action Item

- Design request to send to STP during J-CORES cruise

11. Next meeting

Most DCMG members will meet on the JR during the J-CORES data tool test. Hence the need for another meeting in the fall was deferred until the February-April 2006 time frame.

Appendix A

Ann Klaus' Notes from Data Management Spin Off Meetings 26-28 June 2005

Publications Meeting

Ann Klaus (AK) met with Hans Christian Larsen (HCL) and Manu Soeding (MS) on Sunday, 26 June 2005, and reviewed the following. Tom Janecek (TJ) was also there for part of the discussion.

Proceedings DVD

HCL reported that we were "pretty close" to a final design.

Packaging: Need to make sure that we use high-quality packaging and paper for the insert
Cover Art: Not everyone was happy about the half-tone design, the felt it was hard to read the title through the photo. HCL wants to see this design and the one with solid black on the bottom with several varied photos. Need to send cover samples to HCL.

Make sure type and design of cover is consistent with the art used on the web version of the publication; HCL commented that he liked the font but wondered if it might look out dated in a few years (because he felt it looked trendy)

Insert: Add short instruction to describe CD contents and operation (for example, "To view volume, insert disc and open "contents.pdf")

Disc: Preferred the half-tone over full color

Release of Proceedings Volumes

HCL and TJ decided the *Proceedings* volumes should not be released until the end of the moratorium period.

Exp. 302 Proceedings Front Matter

Member page for ECORD: Correct spelling of Murchison House

Participants' pages for ESO:

Limit participants list in the Proceedings to include those that contributed to the volume contents. List in the following categories:

Expedition Science Party

Expedition Participants (for MSP, this includes onshore and platform)

Other Scientists

(List both groups in alpha order, by last name; include title and contact information)

IO Support Personnel and Subcontractors (get revised wording)

Captain

Drilling Superintendent

Technical staff supporting labs

National Observers

(List in alpha order, by last name; include title and affiliation, not address or email)

Other notes:

Include a complete list of everyone who participated in carrying out the seagoing and onshore portions of the expedition in the Preliminary Report)

WH will provide list of technical staff for Exp. 302 onshore party

Exp. 302 VCDs

HCL and Dan Evans agree that we should get the raw data and format the VCDs to "ODP" style for the publication. Dan noted he never intended the layout we received to be the final layout used for publication.

Synthesis

HCL will write a description of the synthesis requirement and circulate it to us before final publication.

HCL and TJ feel this should be included in the Co-Chief agreement. (HCL plans to draft a list of report/publication requirements for the Co-chief agreement).

Index

HCL and Frank Rack were both supportive of the idea to use some funds that were originally budgeted for preparing indexes and printing books to research the best way to electronically index the *Proceedings* content. The concept was also mentioned to Bernard Miville briefly.

ACTION: The USIO will develop a plan and discuss it with Frank to determine if we'd need a Program Plan change (HCL and TJ didn't think we would).

Publication Submission Procedures

We discussed the process used in ODP (and now) for handling manuscript submission and peer-review. HCL and MS both said that the process is very similar to that used for IODP drilling proposal submission and review and that they have contracted with a programming company to develop a database for the proposals. There is a chance that if not all of the funds are used for this project, IODP-MI might be able to get the same firm to modify the database parameters to that it could also be used for publications (and possibly also sample requests). However, this would require approval by Manik Talwani. Since the work on the proposal database won't be completed until near the end of the fiscal year, it's unlikely we'll know if this is really possible until late in the fiscal year.

We plan to include publication submission guidelines in the Sample, Data, and Obligations Policy as an appendix. We will use the draft HCL sent to the IOs in early June as a starting point.

Proceedings Publications in Phase 2

HCL said that regarding production of the Proceedings volumes in Phase 2 there are probably three possible scenarios for the IOs: (1) IODP-MI instructs the USIO do carryout the same work as in Phase 1 and CDEX to produce Chiky publications, (2) IODP-MI instructs one IO to produce volumes for all IOs, or (3) IODP-MI issues an RFP for production of all volumes and IOs and external vendors compete for the contract. It is expected that in any of these scenarios, each IO will be responsible for preparing publications from expeditions up to a certain stage including scientific editing, i.e., to the stage of the first post cruise meeting.

Possibility of an IODP Journal

HCL said that his goal for FY06 is for IODP-MI to take the lead (with support from the IOs) on developing agreements with the leading publishers of IODP work to provide the IODP science community with open access to articles published in commercial journals. He said he plans to do this, and observe how successful the community is at publishing program-related results in journals when the SR isn't an option for science papers before IODP-MI might pursue a new peer-reviewed and open access journal accepting papers based on scientific drilling and borehole observatory science.

Address Database

Add MS to the TAMU change of address email distribution so that he can update the IODP-MI distribution list. There might be a need for an advanced and programwide address and contact data base.

Sample, Data, and Obligation Policy Review Meetings

1. Ann Klaus (AK) met with Hans Christian Larsen (HCL) on Sunday and reviewed all USIO comments from May 2005 draft.
2. AK, HCL, Tom Janecek, John Firth (JF), Kuro Kuroki, and Walter Hale reviewed an updated version of the document on Tuesday.

ACTION: AK will revise the document and after review by JF, circulate it to IODP-MI and IO representatives from curation and publications.

Curation/Data Management Breakout Session

Ann Klaus (AK), Hans Christian Larsen (HCL), John Firth (JF), Kuro Kuroki (KK), and Walter Hale (WH) met on Monday and Tuesday and discussed the following. Tom Janecek (TJ) was also there for part of the discussion.

Single web-based sample request

Everyone supported the plan to create a single sample request system that is used by curators from all IOs and at all repositories. A database system is needed that will facilitate logging all sample requests and associated curatorial information that is accessible from all repository locations (shipboard and shore-based).

ACTION: For the immediate short-term, requesters will access the riser and MSP sample request forms from the IODP-MI web page.

ACTION: During the coming months, Curators will develop a single sample request form (based on existing forms with modifications to make it suitable for all platforms).

ACTION: Hopefully by the end of summer, IODP-MI will know if FY05 funds to initiate the development of a database to facilitate management of sample requests and associated curatorial information are available. Limited funds for data IODP-MI data management have been requested in the FY06 APP.

IODP Chief Curator Model

In ODP, one Curator was responsible for establishing sampling plans and approving sample requests during moratorium, and the Repository Superintendents handled postmoratorium sampling requests under the oversight of the Curator (if requests were straight forward they could approve; if they needed guidance they would go to Curator).

In cases where a sample spanned more than one repository, each Repository Superintendent would handle the part that related to their part of the collection and any questions or problems would be communicated between all the repository staff.

Now that we have three Curators, we need to establish new operational and authority guidelines. The group discussed the concept of having a Chief Curator who reviewed all sample requests and decided against this. A document was drafted to outline operational and authority guidelines for IODP. This included descriptions of the roles and responsibilities of each curator related to the platforms and the repositories. (See bottom of file.)

During the discussion, KK announced that he has recently been assigned to serve as Curator for the KCR (in addition to serving as Lab Officer that sails on an A/B rotation) and Kazuho Fujine will be a curatorial assistant. He also noted that the Japanese have not developed a curation policy yet, but he believed that JAMSTEC may require special mechanisms for monitoring sample requests for riser core, regardless of where the core resides. HCL said in principle, all IODP cores will have to follow same procedure and be accessible on the same conditions and urged JAMSTEC/CDEX to articulate any special requirements as soon as possible for the consideration by IODP-MI.

WH explained that Ursula will serve on the SAC for MSP expeditions, approve moratorium sample requests, and serve curator role for onshore party, and WH will be responsible for approving postmoratorium requests and overseeing the processing of all on-shore party sample requests.

Moratorium

The group discussed the definition of the "moratorium period" and came up with the following definition:

The moratorium period begins either after the conclusion of a cruise if the majority of the sampling occurred during the cruise, or after the conclusion of the expedition onshore sampling party, and is one year long.

DSDP and ODP Core Redistribution Plan

TJ explained that the Lead Agencies had IODP-MI remove the USIO funds for this project from the FY06 budget in order to meet the \$21M budget target. IODP-MI plans to submit the cost as an extra project and try to defend starting the project in FY06. The USIO budget submission only covered USIO costs and not the cost to buy and build KCR racks, which will be needed before any core can be sent to Japan. We discussed that if the project was funded in FY06, we could initiate the transportation of core between the BCR/GCR and the other U.S. repositories since BCR has all their racks installed.

Database Ideas

As discussed above, the Curators all agreed there is an urgent need for all Curators to have access to all core sample request data and were looking forward to hearing what recommendations the data management staff came up with to manage curatorial data.

During this discussion HCL mentioned that he thinks there needs to be a requirement that DIS (which is a tool not a database) only be used shipboard and that the data be migrated into Janus or J-CORES after shipboard operations. The data management group discussed this concept during their meeting.

Joint Expeditions

In an ongoing discussion that began in the IODP-MI Publications Task Force, as requested by HCL, the USIO provided a summary of their understanding of what elements will be joint and separate related to scheduled joint expeditions (see table below). The USIO asked for clarification as to handle certain aspects of 303/306 and 309/312. In addition, the group discussed the need to develop a standard procedure for determining what would be joint and what would be separate on future joint expeditions prior to the expedition initiation. HCL also discussed the issue of 303/306 with Mike Coffin (Co-chair of SPC):

Joint Expedition Summary - USIO understanding as presented at the meeting

	304/305	303/306	309/312
Scientific Prospectus:	Joint	Joint	Joint
Preliminary Report:	Separate	Separate	Separate
First postcruise meeting:	Joint	Separate	Not defined ⁶
Second postcruise meeting:	Joint	²	Not defined ⁷
Sampling meetings:	Joint	Separate ³	Not defined ⁸
Moratorium:	Not defined ¹	Not determined	Not defined
Proceedings volume:	Joint	Separate ⁴	Not defined
Science Party:	Joint	Joint ⁵	Joint

¹ Per interim policy 12 months after release of samples or data. This will probably need some resolution as sampling party was held in early June but people will not receive their samples for several more weeks.

² CC want separate; SS advocated for joint meeting

³ Because of volume of samples and length of time between cruises; 304 was scheduled 4 weeks after Exp. 306 ended to be able to accommodate requests from Exp. 306 scientists.

⁴ Separate but I think that these should be linked/organized together on the web

⁵ This is a minimalist example

⁶ If there is to be a single *Proceedings* volume, there should only be one postcruise editorial meeting, but I do not know if the single volume concept has been adopted.

⁷ It makes sense to have a single meeting

⁸ Exp. 309 will need access to Exp. 312 cores as soon as possible after Exp. 312, so many may need to travel to repository to sample and this should be coordinated. Exp. 312 may have the opportunity to sample Exp. 309 cores at sea, but doubt there will be time, so they will probably need a sampling party as well.

Notes:

Mike Coffin stated that the Expeditions 303 and 306 were designed as a joint expedition that stemmed from one proposal and should remain so. It has subsequently been determined that 309/312 also will be a joint expedition.

HCL and TJ agreed that for future expeditions, by the conclusion of the preexpedition planning meeting, the following should be determined: dates of moratorium period, potential need for shore-based sampling party. Decisions on joint expeditions (single or joint science party, sampling parties, publications, and postcruise meetings) should be made during scheduling based on SPC advice.

+++++

DRAFT Curation Roles and Responsibilities

Curator Roles and Responsibilities

The IODP repositories are overseen by three IODP Curators:

Kochi Core Repository: Kazushi "Kuro" Kuroki

Bremen Core Repository: Dr. Ursula Röhl

Gulf Coast Repository, West Coast Repository, and East Coast Repository: Dr. John Firth

Each Curator plays the roles of "Platform Curator" and "Repository Curator." In addition, JF will coordinate the handling of education/museum requests.

The Platform Curator is responsible for overseeing all curation tasks from the pre-planning stage through arrival of core materials at the residing repository. From the end of the expedition through the moratorium period, the Platform Curator also retains responsibility for use of the core. The Repository Curator is responsible for preservation of the core once it arrives at the repository and use of the core after the moratorium period ends.

To conduct the IODP geographical distribution model successfully requires three curators with equal authority and sufficient expertise and knowledge about each expedition (from expedition planning and budgeting through acquisition of the core and planning/execution of moratorium usage) and the archive collection.

IO Responsibilities Relating to Curation Activities

The platform Implementing Organization (IO) is responsible for the costs and assurance that the following work is completed:

Planning the procedures and costs associated with recovering core

Development of Sampling Strategy

Collection of cores

Minimum measurements of cores

Shipboard curation

Sampling on drilling vessel

Shipment of cores to the residing repository (i.e., final location of core materials)

Shipment of samples from vessel to participants

Oversight of sampling during the moratorium period

Transportation costs associated with support of platform science. This may include shipment of old cores from any repository to/from the IO drilling platform.

The residing repository is responsible for costs associated with:

Moratorium sampling at the residing repository

Postmoratorium sampling

All other postmoratorium curation activities

Curatorial Planning

Curatorial planning should begin when exhibition budgets are initially prepared ("Project A's") for review/ranking by Science Planning Committee/Operations Task Force. Once an expedition is scheduled, refined planning should be led by the Expedition Project Manager (Staff Scientist) and Platform Curator, in consultation with the Co-Chiefs and other operator staff, as well as the Repository Curator when applicable. This planning, which will result in the creation of the Sampling Strategy, should take place at the preexpedition meeting.

Through the creation of the Sampling Strategy, the Platform Curator should work with the Repository Curator, Expedition Project Manager (Staff Scientist), and Co-Chiefs to develop a plan for how any shore-based work (measurements or sampling) will be accomplished.

By the conclusion of the preexpedition planning meeting, the following should be determined: dates of moratorium period, potential need for shore-based sampling party, and issues related to joint expeditions (single or joint science party, sampling parties, reports and publications, and postcruise meetings).

If deviations in minimum measurements protocols are identified or anticipated, the outcome of any planning exercises should then be communicated broadly so that IODP-MI, the other repositories, and the Science Advisory Structure (primarily the Science Technology Panel) are all informed at the earliest time possible.

Appendix B

Memorandum

From: IODP Data Management Coordination Group (DMCG)
To: Scientific Technology Panel (STP)
Subject: IODP ("Minimum") Measurements
June 30, 2005

At its June 27-30 meeting in Edinburgh, the IODP DMCG discussed the topic of "IODP Minimum Measurements", which is also an agenda item at the upcoming STP meeting. The DMCG's interests in the definition of IODP measurement deliverables are the creation of an information management system with measurement types and parameter at its core, the generation of metadata for the future IODP data portal, the prioritization of software tool developments for the IODP, and miscellaneous issues associated with the data management requirements of any such definition. The DMCG recognizes that the STP may have additional interests and assumes that STP will ultimately formulate an IODP measurements policy. The DMCG therefore submits the following thoughts to the STP.

A multi-tier scheme to classify all existing and potential future measurement types may be more appropriate than a simple set of "Minimum Measurements" because a number of different aspects need to be taken into consideration. The definitions of measurement types should encompass all measurement types and each measurement type should be associated with one of the tiers.

- The definition of IODP measurement deliverables should be defined based on a complete list of all measurement types, each composed of one or more "parameters". The IOs are in the process of completing such a list

Definitions of measurement deliverables should be relevant to specific platform needs and include identification of resources for procurement, capital replacement, and continued operation of the measurements.

A comprehensive definition of measurement deliverables should be compatible with the future IODP third-party tool policy.

The DMCG thus came up with the following first cut at a list of IODP measurement deliverables:

Tier	Type	Comments
1	Operations and Safety Measurements	Platform-specific measurements required to safely drill a well, recover core, and/or instrument a well. This is a Platform Operating Cost (POC) and selection and implementation of measurement systems and QA/QC is therefore the responsibility of the respective IOs. Examples: Seafloor depth; weight on bit; abundance of volatile hydrocarbons in cores.
2	IODP Minimum Measurements	Measurements to be taken by all IOs on all cores, or, in the case of downhole measurements, at all sites. This is a Science Operating Cost (SOC); the question here is if procurement and capital replacement should also be funded through SOC if a given IO does not have the resources to provide a system, given the fundamental requirement to collect these data. QA/QC definition should be coordinated through the IODP-MI. Examples: Curatorial data; digital image of each core section surface; magnetic susceptibility; downhole natural gamma ray;
3	IO Standard Measurements	Measurements are collected if (1) the measurement system is available to the IO and (2) measurement conflicts and priorities do

Data Management Coordination Group Meeting #2
Edinburgh, Scotland - June 27-30, 2005

		<p>not prevent collection and (3) the type/quality of the recovered material or hole do not prevent the collection of adequate data.</p> <p>Selection, implementation, and capital replacement of these measurement systems are POCs. Operation costs are SOCs. QA/QC definition is the responsibility of the respective IO but should be based on emerging IODP standards whenever possible.</p> <p>Examples: Igneous rock thin section descriptions; abundance of pore water constituents; microbiological contamination tests.</p>
4	Third Party Measurements on IO Facilities	<p>Measurements with systems provided by third parties. Data and report must be made available to IODP.</p> <p>Deployments are according to Third Party Policy and funding is through third party.</p> <p>Example: A "miniprobe" to estimate elemental abundance in thin sections; new tool to measure downhole pressure.</p>
5	Other Measurements	<p>Any other measurement taken on IODP core material with external funding at any facility.</p> <p>These measurements are externally funded and IODP assumes no control or responsibility.</p>