

Minutes of IPY Data Subcommittee Meeting

2nd March 2006

British Antarctic Survey, Cambridge

Purpose	1
Attendees	1
Minutes	2
Introduction from Dave Carlson	2
Committee Operational Procedures	2
How do we operate?	2
Links to the JC	2
Meetings	2
Funding for the Data Subcommittee	3
National funding for Data Management	3
What is our decision making process?	4
Procedures for modifying membership	4
How should the JC and the data subcommittee work together?	4
Communicating with other IPY Subcommittees	4
Use of a web site	5
Data Policy	5
Data Access	6
How soon should data be made available?	6
Metadata	6
Summary	7
Terms of Reference (ToR) of the DMSC:	7
General	8

Purpose

The purpose of the meeting was to bring together the IPY Data Subcommittee for the first time, and establish operational procedures, Terms of Reference, finalise the data policy and data strategy, and determine what to present to the wider data management workshop to be held on 3rd and 4th March 2006.

Attendees

Taco de Bruin (Co-chair) - Netherlands Institute for Sea Research, Netherlands

Mark Parsons (Co-chair) – National Snow and Ice Data Centre, USA

Ray Harris – University College London, UK

Ellsworth Le Drew – University of Waterloo, Canada

Eberhard Fahrbach (IPY JC) – Alfred Wegener Institute, Germany

Hannes Grobe – Alfred Wegener Institute, Germany

Hakan Olsson – Forest Resource Management and Geomatics, Sweden

Joan Eamer – UNEP

Dave Carlson – IPY IPO

Evgeny Vyazilov – All-Russian Research Institute of Hydrometeorological

Information – World Data Centre, Russia

Vladimir Papitashvili – University of Michigan, USA

Kim Finney (for Nathan Bindoff) – Australian Antarctic Data Centre, Australia

Minutes

Introduction from Dave Carlson

Dave Carlson talked about the progress that is being made for the IPY, and that the data produced during IPY may be the most important legacy for the next 50 years. The IPY community includes an unprecedented depth of investigations that were not present in the IGY, and incorporate diverse components such as archaeology, palaeo-ecology, linguistics etc. Not all these communities have the same experience of data management, and therefore there is a need for data management outreach across all the communities contributing to IPY.

Committee Operational Procedures

How do we operate?

Links to the JC

- At the working level the Data Management Subcommittee (DMSC) should make decisions, and provide the JC with results/solutions, as the DMSC are the expert group. Decisions should only be referred to the JC when the decisions are associated with high level partners (Eberhard Fahrback)

Meetings

- It was recognised that it will be hard to get all the members of the DMSC together again due to funding and existing busy schedules of the members
- Dave Carlson said that if the DMSC think they need to meet every 6 months, then they should inform the JC, so that the JC can try to find the necessary funding.
The group agreed that they should aim to meet every 6 months
- The use of teleconferencing was also discussed. The IPO has a site licence which the subcommittee can use, but the option of using some freeware was also mentioned. Microsoft Live Meeting was also mentioned
- The use of group web email facilities was discussed, which allow creation of threads to enable the whole group to read and follow discussions in the correct order.
- Eberhard Fahrback pointed out that teleconferencing can be difficult for non-english speakers, and that meetings are time-consuming, and therefore he prefers the use of group email facilities. However the group agreed that due to the short timescale before IPY, it would be good to use teleconferences to force people to respond straight away, and come up with recommendations. **It was agreed that regular teleconferences would be held, with formal agendas**

Actions:

- 1. Taco and Mark to set up a web email facility for the committee**
- 2. A telecon will be held every 2 months – Mark to organise?**
- 3. An ad-hoc meeting will be held at the SCAR meeting in Hobart for the members of the committee who are there – Taco will organise this. Dave Carlson suggested that this could be tied in with the IPY discussion meeting being held on 8th. It was suggested that the committee should use**

this opportunity to present to the science community about what is being done for IPY data management

- 4. Taco and Mark to try to organise 6 monthly meetings for the committee (the meeting in Hobart does not count as only a small proportion of the committee will be there)**
- 5. Mark to set up a collaborative wiki for producing, sharing and editing documents such as the IPY data policy and strategy – e.g. Google groups**

Funding for the Data Subcommittee

- **Taco proposed that the DMSC should try to become a CODATA task force, as this may bring in some extra funding**
- **Vladimir said that he would try to get some extra funding from SCAR for the DMSC**
- **It was recommended that all members of the group should look for further sources of funding for the DMSC**
- **It was agreed that the DMSC should ask the IPO to send a letter to the funding agencies of each member of the DMSC asking them to support the DMSC member with the important work that they are doing (both financially and by giving them time to contribute to the work of the DMSC)**

National funding for Data Management

- It was pointed out that there will be a need for national funding agencies to set aside some money for data management work during IPY. Some national funding agencies are in the process of deciding what should be spent on IPY, and therefore need to know very soon about what they will need to set-aside for data management (e.g. Sweden).
- The need to use existing data management structures and facilities for IPY was mentioned, and it was highlighted that these existing structures need to be made aware of what else will be expected from them for IPY. It was suggested that a big meeting could be held in 2007 in order to identify what everyone is already doing.
- **It was decided that the DMSC should ask the JC to send a letter (separate to the one about the subcommittee members), to the national committees and funding agencies asking them to address data management issues.** This letter should lay out the specific components of good data management – Ray Harris identified a section in the ICSU PAA on professional data management, in which it is stated that all projects should include, and be evaluated on data management, which could provide the basis of the letter. The letter should also refer to existing national data policies, and the need to integrate data management for IPY with existing national structures and systems for data management

Actions:

- 1. Dave Carlson agreed to produce the letter, and send it out, but asked to be given information about who it is targeted at, and what the main content would be (he advised that the letter should contain 3 main bullet points). The DMSC agreed that the main targets for the letter would be the IPY national funding agencies, ICSU and the WDCs. Mark and Ray agreed to work on the main content of the letter**

What is our decision making process?

- In general the group will aim to agree by consensus, and will only pass decisions up to the JC if someone adamantly disagrees with a decision being made by the DMSC.
- When an important, time-dependent email is sent out which requires a response, it will state how soon a reply is expected, and if no response is received, then this will be taken as agreeing with the email.
- When replying to emails, the 'reply-all' option should not be used, as it causes a lot of extra work for each committee member, and can bias responses from other people

Procedures for modifying membership

- If one member leaves, then the DMSC should seek approval from the JC to invite another person from their institute to join.
- The committee will invite selected professionals to join in the meetings when they are needed
- If a gap in expertise is noticed, then the DMSC should suggest to the JC that a new person is needed – the JC have said that they do not want the subcommittees to grow too big
- Members of the DMSC are primarily members due to their expertise, before their national interests
- It was recommended that members from experienced groups such as IODE and GBIF might be beneficial, but the group agreed that this should only be reviewed if it is noted that there are problems with the current membership

How should the JC and the data subcommittee work together?

- **The next meeting of the JC is 20th, 21st, and 22nd April – the report from this meeting should be ready for then**
- **Any decisions made by the DMSC should be sent to the JC Chairs with Eberhard and Dave Carlson copied in.**
- The IPO is setting up a Project Coordinators Distribution list, which will be used to provide factual information about what is happening for IPY. **It was recommended that the committee should use this as a communication channel for data management activities.**
- Dave Carlson suggested that the IPO emails which are currently sent to the JC members could also be sent to the DMSC members
- **It was suggested that the JC meeting agendas should formally include a data management section, and it was suggested that one of the DMSC Co-Chairs should attend these meetings??**

Communicating with other IPY Subcommittees

- There are two other official subcommittees for IPY – Education and Outreach, and Observations.
- Education and Outreach – Mark suggested that there is unlikely to be need for a close linkage between the DMSC and the Education and Outreach committee, although the DMSC will be happy to answer any requests for information from them

- Observations – there is likely to be a lot of needs for the DMSC and the Observations Subcommittee (OSC) to work closely – Mark has tried contacting them, but with little success.
 - The difference between the OSC and the DMSC is that the OSC will be responsible for ensuring that observations are done in a coordinated way, whereas the DMSC will be responsible for ensuring that data collected from IPY observations are managed in accordance with the data policy.
 - **The OSC are meeting in Potsdam on 27th and 28th March, and this should be used as an opportunity for the DMSC to let them know that we want to work with them, and update them on what we are doing.**
 - An initial project for the OSC and DMSC would be to assess data flow for the IPY by carrying out a ‘Data Questionnaire’. The IPO will be interested in the results of this, as it will help them to identify which of the proposed projects are definitely funded and going ahead.
- It was agreed that it would be useful to know what is going on in the other IPY subcommittees – it was suggested that the use of wikis hosted on the main IPY site would mean that each subcommittee could keep updated with news.

Use of a web site

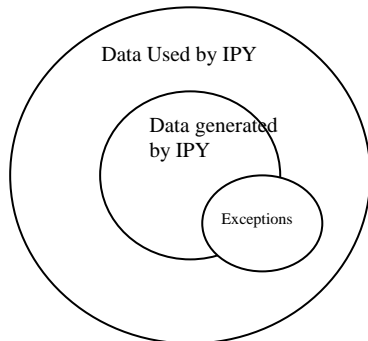
- The group agreed that all official information from the DMSC should go through the official IPY site – to give the DMSC an official presence.
- The DMSC also want to use the web for conversational material, and unofficial material, which will need to be password protected. **Mark will discuss this with Dave Carlson – Joan offered to assist with this.**

Data Policy

- The data policy needs to be focussed around the aims of IPY and how data management can help achieve these. The IPY framework document can be used as an outline for this.
- The data policy will be about the rules which should be applied to data management for IPY, whilst the strategy will detail how this will be implemented and enforced – the policy should be kept separate from the enforcement.
- The question of whether to include publications in the policy was discussed, as ICSU see data and information as a continuum that leads right through to publications, but it was suggested that opening up the problem of publications would not be good.
- Enforcing the IPY Data Policy can be done through encouraging national funding agencies to withhold some funding, and/or to rescind the IPY blessing if a project does not conform to the policy.
- As the policy is now being updated, it will be necessary to ask all projects to re-accept the new policy.

Data Access

- How do we define IPY data?



Data Used by IPY – **need to respect the data policy of the supplier**

Exceptions – **may be ethical (data related to locations of endangered species), scientific (data which needs more processing), commercial**

- The general policy is free and open access to data, but the policy needs to recognise that there are certain limitations in certain areas, which is shown in the above diagram by the 'Exceptions' box. However, the exceptions are only related to access to data – ALL data should be preserved as the legacy of IPY.
- Free and Open Access will need to be defined in the Policy – use p24 of the ICSU PAA – it would be good to reference the ICSU and WMO statements on this.

How soon should data be made available?

- There needs to be a distinction between observational (routine) data which should be made available very soon after collection, and data collected for a specific project/publication which should be made available after a specific time following data collection/dataset completion– e.g. 6 months. However complications can arise from this as there is a need to differentiate between collection of a sample, and completion of a dataset
- But, setting more, tighter limits may just alienate the science community further, and make them think that data managers are not in-touch with the science world – but the policy should still require prompt publication after dataset completion

Metadata

- Metadata should be available for everything that is produced for IPY, including annual reports etc
- Pre – project metadata would be useful for data managers to know what to expect, and may also be useful for scientists to know more detail about what is going to happen during IPY – this could help collaborations between projects which is an important component of IPY. Taco mentioned that Cruise plans are used a lot in the Oceanographic community for scientists to plan who to work with. However it should not be called metadata, as metadata should be linked closely with data to make it really useful, and planned work is not data.
- **The DMSC agreed that metadata must be submitted together with all IPY data, and that pre-experiment data descriptions (which should be very short and uncomplicated but include Title, High Level Parameters, Time Frame, and Location) should also be submitted.**

- IPY needs to be visionary and idealistic – where do we want data management to be in the future? Recognition of data publishers is vitally important – the policy needs to include a requirement for formal acknowledgement of data publishers, and an ability to track usage of data, so that we can identify the most useful/used datasets. In social sciences this has an added dimension when using questionnaires – need to recognise ownership of the actual questionnaire (structure, questions etc), and ownership of information gained from the questionnaire – i.e. from local communities
- The DMSC need to seriously consider the Arctic Communities, and their needs and rights to see the data – not just the social data, but also the physical data about their environment.

Summary

The data policy needs to include an introduction stating why we have a data policy for IPY, and how the objectives of the data policy relate to the objectives of IPY. The Policy needs to include:

- Free and Open Access – Joan suggested a good statement about open access which recognises the need for certain well-defined restrictions
- Acknowledgement of scientists who produce and publish data, and owners of knowledge (e.g. local populations) for use of data
- Outreach to local communities to enable them to access data about them and their environment
- Metadata for all data (regardless of whether it has restricted access etc)

The issue of IPR for indigenous populations needs to be specifically addressed in the data policy – should this be done by the DMSC or the OSC? Advice could be sought from the Convention on Biological Diversity which has an ‘Indigenous Working Group’. Joan and Berger will work on this.

Actions:

- **Mark will put the current policy on a wiki so that everyone can get access to it, and suggest how it should be amended**
- **There will be a Telecon in 1 month’s time so that the Data Policy can be approved in time for the JC meeting in April (it will need to be done before the 15th April so that it can be sent out in time for the meeting)**

Terms of Reference (ToR) of the DMSC:

- The DIS is a servant to the DMSC – the DMSC provides the guidance, and then the DIS implements it
- The DIS and DMSC must provide clear information to the whole IPY community to ensure that everyone knows clearly what is expected of them
- The DMSC is responsible for defining data management requirements for the IPY – there will be a need to consider international projects, and who will be responsible for data management in these projects.
- The DIS at the moment is a loose group of data centres and interested parties which requires a clear policy defining how interoperability, long-term storage etc will be achieved
- The ToRs will need to evolve as projects and the DIS evolve – we need to accept some ambiguity at this stage.

- There was concern that although the DMSC may agree to their ToRs, it may be difficult for them to achieve them due to a lack of time and resources

Action:

- **Each member of the DMSC should review the ToRs which were handed out at the meeting, and send any suggested changes to the group.**

General

- Vladimir recommended that there is also a need to address the role of the existing WDC structure for IPY and the national funding of these.
- Dave Carlson's vision is that everything being done in IPY should be viewable on a system like Google Earth, before, during and after the IPY – we must remember that we have an obligation to make information available in a very timely fashion, and get it right, as we will have a high public profile during the IPY.

Helen Campbell
20th March 2006

Minutes of IPY Data Subcommittee Meeting

4th March 2006

County Arms Pub, Cambridge

Purpose

The purpose of the meeting was to agree on the immediate actions to be taken by the IPY Data subcommittee, following the IPY data management meetings held in Cambridge from 2nd to 4th March 2006.

Attendees

Mark Parsons, Taco de Bruin, Joan Eamer, Kim Finney, Ellsworth Le Drew, Håkan Olsson, Birger Poppel

Minutes

1. Data Management Letter

A Letter needs to be sent to IPY funding agencies about data management as soon as possible. This needs to:

- Explain why data management is needed – emphasise the unique character and multi-disciplinary nature of IPY, and that data management is fundamental to ‘preserve the unprecedented quantity, quality and variety of data which will be produced in this programme, which uniquely combines data produced from research in all areas of science, and places a strong emphasis on the human dimensions of all the research that is being done.....’ The quote on data legacy from the IPY framework document could be used. Also emphasise that good data management will facilitate new links being forged between scientists from previously disparate communities. Finally, make the point that by being involved in data management, they will be directly contributing to the massive legacy which the IPY will leave for the next 50years....
- Encourage coordination of data management within their countries, and to make them aware that their data management will need to be coordinated at both a national and international level for IPY
- Let them know that the Data Management Subcommittee has been established, and that we are there to help national organisations with their data management. Let them know that we will contact them again in the future as data management requirements for the IPY develop – ‘we look forward to working with you, and will be in contact again soon with further details of data management activities for IPY’
- Encourage them to think about how their data will be preserved in the long-term, and suggest that they contact suitable data archiving facilities (either national or international), to discuss the data that will be produced by their nation, and what will need to be done to ensure long-term preservation for future generations
- Recommend a data manager being appointed for each IPY project
- Set out the minimum standards for data management for IPY projects, which includes ensuring that each project considers data management during the project application stage, and that money for data management for each project is identified and set-aside at the start. This should also include pointing them to the ICSU Data Management PAA, and highlighting the

component, which states that data management should be carried out in a professional manner and that money should be set-aside for data management for all science projects. It must also be emphasised that IPY data management will require additional work beyond the systems that already exist for data management in their country

- **Explain that fair, equitable and open access to data is key to advancing human society's understanding of our planet, and is essential to attract young researchers, and inspire young people to get involved in science.**

2. IPY data coordination person

It is vital to kick-start IPY data management coordination, in advance of funding for the DIS. It was agreed that employment of a temporary coordinator was more vital than funding travel for the data management sub-committee, at this stage. The post would initially be for 6-12months, and would be focussed on going through the IPY project honeycomb diagram to identify major areas of data management, and producing a pre-collection data survey. This would be useful in order to:

- **Get a clearer understanding of what data will be coming out of the IPY projects, in order to identify data management work which will need to be done, and any possible gaps in current data management structures/facilities**
- **Get a clearer understanding of areas where there may be gaps in data collection**
- **Provide information for scientists on overlaps/gaps in data collection, and potential partnerships based on data collection.**

Action: IPY data sub-committee to recommend to the JC, that they contact WMO and ICSU to request temporary funds for this post

3. User-friendly version of ICSU PAA

It was agreed that it would be useful to produce a more user-friendly version of the ICSU PAA, and how it relates to IPY. The document should start with a vision of what we want the data management product to be – for example – facilitating data from disparate science disciplines being integrated and compared, to lead to new scientific discoveries, and to provide new unique educational resources at schools.

Action : An Outreach Task Force be established consisting of Taco, Ellsworth, Joan and Ray to produce the document, and the Data subcommittee to ask ICSU/IPO/NSIDC to turn it into a professional distributable document.

4. IPY data management architecture

The IPY Subcommittee will need to provide the initiative for data centres to work together. They need to set out a roadmap (inc. data architecture), of where we are trying to get to with data management, and then ask the data centres what they can do to assist with this process.

5. Other

- **The Observations Subcommittee should be encouraged to send a similar letter to that described in point 1, to observational agencies involved in IPY**

- **Need to get WMO as involved in data management for IPY as ICSU – Taco will contact Ed Sarukhanian**
- **GEOSS – it was agreed that adopting the principles laid out by GEOSS would be beneficial for IPY. IPY could form a ‘Community of Practice’ for GEOSS. Ellsworth will informally suggest this at the GEOSS meeting in Rome on 8/9th March, and report back to the Data Subcommittee. If this is all agreed, then the data sub-committee will recommend to the JC that this is a good approach.**
- **The IPY data policy must be finalised by the next JC meeting**
- **An IPY Data Sub-committee meeting will be held by telecon end Mar/beginning Apr.**

**Helen Campbell
9th March 2006**