International Polar Year 2007-2008: Data Policy

DRAFT - JUNE 2005

Introduction

The IPY 2007-2008, an intense, interdisciplinary, and internationally coordinated campaign of research and observations, will deepen understanding of polar processes and their global linkages. IPY knowledge and the observations upon which it is built must be effectively managed to ensure the greatest benefit in the future. IPY-generated data should be carefully and thoughtfully collected, used collaboratively, and adequately preserved. IPY data should be freely available, although in some cases restricted access may initially exist.

IPY data will be multidisciplinary and disparate. This policy aims to provide a framework for these data to be handled in a consistent manner, and to strike a balance between the rights of investigators, the rights of indigenous peoples, and the need for widespread access through the free and unrestricted sharing and exchange of both data and metadata. The policy must also be compatible with the data principles of the sponsoring bodies, ICSU and WMO, and other relevant international agencies. The IPY data principles described here draw heavily on those of WCRP CLIVAR.

Since IPY 2007-2008 is a distributed programme, implemented through a number of self-managed IPY Projects, the principles enshrined in the Data Policy must be applied to data in each Project. In order to be considered as officially IPY, each Project must follow the IPY 2007-2008 data and information management policy, including submission of metadata and data according to an agreed timetable, and must include an appropriately funded data management plan. In a similar fashion, in order to take part in an IPY 2007-2008 Project, participants must agree to submit information and data from their component of the project, and comply with the IPY 2007-2008 Data Policy.

Definitions used in the IPY Data Policy

1. IPY data
"IPY Data" consist of directly observed data, derived data, gridded fields, model outputs and other data products generated and/or used within activities that are undertaken as part of endorsed IPY 2007-2008 Projects. Other data and related products that are relevant to the objectives of IPY Projects, and are generated during the IPY time-frame are also valuable, and IPY should strive to ensure that all data relevant to IPY are accessible freely and without restriction, including those collected by other projects and programmes.

2. Metadata
Metadata are defined as the descriptive information such as content, quality, condition that characterizes a set of measurements. Metadata allow someone looking for a

particular type of data to find its location, and know whether it is appropriate for a particular use. They usually describe who measured what parameters, where and when, how, and who to contact to obtain the data. Metadata should be the principal vehicle for documenting known data quality Metadata must be in a searchable database. Much information on what sort of data will be collected is known when a project is originally proposed, and this is described as 'catalogue' metadata.

IPY Data Policy and Principles

1. Traditional knowledge and cultural heritage

All IPY participants will respect and safeguard current and traditional knowledge from and about circum-Arctic peoples and all related tangible and intangible cultural heritage.

2. Free and unrestricted exchange

All IPY data should be made available freely and without restriction. "Freely" means at no more than the cost of reproduction and delivery, without charge for the data itself. "Without restriction" means without discrimination against, for example, individuals, research groups, or nationality. For some social science data some restrictions may be necessary for ethical reasons. In exceptional circumstances involving highly specialized or experimental data, principal investigators may temporarily limit access until such time as the data can be adequately validated.

2. Timely exchange

IPY investigators should make data available voluntarily and with minimal delay to maximize their value to IPY. In many cases, particularly for projects making widespread and synoptic observations, data should be submitted in real or near-real time. In cases where extensive post-processing of delayed mode data is needed before a final research quality data set can be generated, early release of a preliminary version of the data is required.

3. Quality control

IPY investigators retain the primary responsibility for the quality of the data they produce and distribute. Data originators and those generating derived data products are required to ensure that their data meet international quality standards wherever possible.

4. Metadata

All IPY 2007-2008 project leaders or participating scientists must submit 'catalogue metadata' to a central database, describing what data they intend to collect as part of IPY 2007-2008 during the planning stage. More detailed metadata must be submitted to an IPY-identified data centre immediately after data collection. Metadata for IPY data sets will be developed and managed in accordance with international standards.

5. Preservation of data

Long-term survival, integrity, and access to IPY data must be preserved for future generations. Internationally agreed standards should be used for the acquisition, processing, and final archival of data and metadata. Data distributed in real and near-real time should, wherever possible, be replaced in a delayed mode after it has undergone quality control and full documentation.

6. Easy access
IPY encourages the use of the most recent advances in communication to ensure widespread access to data collected under auspices of the programme.

7. Use of existing national and international mechanisms and centres
Where feasible, IPY will use existing facilities such as World Data Centres, other regional or national data centres, and recognized metadata centres for the exchange and storage of data. By building on the data management structure of existing programmes, the effectiveness of the data system will be improved and financial costs minimized. IPY should also be prepared to maximize the full benefits of new data and information technologies and capabilities, and to rethink, re-orient, and substitute for existing structures and bodies where it is necessary to achieve its objectives.

8. Reporting Requirements
Data and metadata should be submitted to recognized data assembly centers as well as to appropriate national and international archival institutions so that the collected information may be safeguarded for future analysis. Inventories of data and related information should be readily accessible and updated as needed on a routine basis.

9. Acknowledgement
An ethical policy for data use should be established, complying with the existing data policies of ICSU and WMO, and building upon other already existing models, with special emphasis in incorporating social science data. As a basic principle, scientists should receive due credit when other parties use their original data. Similarly, funding agencies or scientific institutions need to know that their contribution will be properly recognized by users of the data. Hence appropriate acknowledgement of data providers should be part of the IPY 2007-2008 data management strategy. Along with their metadata or data, scientists submitting data to a data centre should provide clear information about their preference for acknowledgment and/or contact prior to publication (by third parties) of studies using those data. Data centres should ensure that this information is distributed with any data when they are released.