

POGO DATA MANAGEMENT PRINCIPLES – DECEMBER 2003

PREAMBLE

The primary objective of data management is to ensure timely, efficient and open access to the best possible data, metadata and associated products, for use and re-use throughout their life-cycle and to prevent loss of data and associated information. There are too many instances where valuable data sets have been lost through accident or neglect, or not been accessible as they are stored in non-digital form or on scientists' own workstations. Moreover, in view of the general sparseness of ocean data sets and the considerable investment required to collect observations, it is critical that data are managed in a way that maximises their utility and impact and thus their benefit to the community.

From the perspective of POGO partners, good data management can be facilitated through adoption of organised and effective practice at the institutional level and/or through partnership with centres who have expertise and experience in ocean data management. There are many different approaches that can be used in such partnerships, ranging from routine data submission through to a more proactive involvement by the data centre in an institute's data management activities. The partnership of scientific agencies and data managers is a valuable one, freeing scientists to carry out research and enhancing their own information sources and systems, whilst preserving the data sets for future use, and enabling research cooperation that might otherwise have been inhibited.

The principles set out below are intended to be “down to earth” and illustrate best practices without considering any technical issues. It is necessary to address the fundamentals first. For example, not all data are held digitally, much data exist in an undocumented form without all the needed metadata, and individual institutions often do not have a clear overview of their data collections. The principles presented here address the fundamental principles rather than the technical modalities of data management and distribution.

POLICY AND STRATEGY

1 POGO Institutes should develop and implement a policy governing data management and data exchange

Oceanographic data are often irreplaceable and are unique. They are also expensive to collect. It is therefore beneficial for POGO institutes to develop their own data policies to make clear the ownership and custody of data, the responsibilities and obligations of those collecting and holding data, data access, data charging and other institutional, legal and contractual obligations. Policies should also promote data sharing/data re-use, both within an institution and among the research community in general, and prevent the loss of data other than by deliberate planned disposal.

There are examples of such policies available both for projects (e.g. WOCE, JGOFS, Argo, etc.), major programmes (e.g. GOOS [www.ioc.unesco.org/goos/GOOSdm_final.pdf], GCOS), and for organisations (e.g. UK NERC [www.nerc.ac.uk/data/policy.shtml]). In addition, a joint IGBP/SCOR policy is under development. Note that the IOC Data Exchange Policy (see 3 below) specifically addresses non-governmental ocean agency data exchange. Several POGO members have sound data policies (e.g. CSIRO), that can be used by other POGO member institutions as a basis for developing their own data policies.

2 Institutional data management policy and that covering cooperation among POGO institutes should recognize and exploit the adopted international IOC Data Exchange Policy and promote its implementation

At the 22nd IOC Assembly, June 2003, the IOC Data Exchange Policy was approved. This policy promotes timely, free and unrestricted access to oceanographic data and associated metadata, and aims to maximise the amount of data exchanged without infringing the rights of data originators. Clauses 2 and 3 in particular address arrangements relevant to POGO institutes. It also promotes the use of national oceanographic data centres (NODCs) as long-term repositories for data and metadata, and encourages capacity building. The IOC Data Exchange Policy is included in the Summary Report of the IOC Assembly

[www.ioc.unesco.org/iocms/files/IOC-XXII_3.pdf]. The important task now is for member countries and their organisations to implement the policy. POGO institutes need to be aware of this data policy and aim to implement it.

3 Monitor the effective implementation of data policy

Many countries and many institutes have established data management policies. However, they are not adequately enforced. The POGO Directors need to agree on a set of indicators to allow them to evaluate the adherence to the policy, and include the status of their data management activities in their annual reports.

4 Data management and resourcing of that management should be an explicit consideration in institutional strategic planning, including participation in research experiments

POGO institutes should promote the benefits of data management in their strategic planning and ensure that research activities make proper allowance for management of data, both for the activity itself and for the benefit of scientists in the future. POGO institutes (and scientists) must build adequate funding for data management into their programmes, and include data management specialists in discussions from an early stage. Data archiving, management and dissemination require special attention and the effort needed should not be underestimated. All too often this is allowed to fall off the bottom of the list of priorities, which may subsequently lead to costly data archaeology and rescue efforts or simply loss of the information altogether. In the event that a conscious decision is made to dispose of particular data sets at an institutional level, POGO members could adopt the policy that they would consult other members for interest in the data, prior to any action being taken.

5 Scientists are rewarded for their research and publications

Currently, there is no real benefit for a scientist who is a data collector if he/she makes the data available through a data centre. POGO Directors should consider developing a recognition programme as an incentive for the scientists to follow good data management practices and produce excellent data sets for the user community.

PRACTICAL MEASURES IN INDIVIDUAL INSTITUTES

6 Use existing data centres (e.g. NODCs or specialist data centres)

POGO institutes should encourage the involvement of (national or specialised) data centres early on in their data management planning, and pass data on to them. This should aim to ensure that responsibilities are properly allocated - to avoid duplication of effort, encourage collaboration, the best use of resources and expertise, and the production of enhanced data sets and value-added products. It should also be noted that expertise at data centres is not limited to physical data alone, and solutions exist for biological and chemical data management. Should there be a requirement to adapt the practices of any of the national or specialist data centres to meet POGO's requirements, POGO should try to influence the appropriate authorities to do so. POGO institutes may also have a role of technology transfer to specialist data centres (e.g. to NVODS activities).

7 Do not "re-invent the wheel"

IOC's IODE is the international manifestation of past and present data management activities supporting science and therefore POGO institutes should in the first instance seek arrangements that can exploit/benefit from this experience (it is the parallel to POGO and GOOS). In addition, there are several other international agencies, organisations and projects which are aware of the issues and problems that face oceanographic data management and distribution (e.g. GOOS, JCOMM, IGBP/SCOR, WOCE/CLIVAR, OBIS, ICES, PICES). POGO institutes should be well co-ordinated with these and take time to find out if there is already a solution to their data management issues, or if there is one can be adapted to meet their needs. *[Note: This would benefit from a focal point where information can easily be found (e.g. a link from POGO web-site to IOC/IODE).]*

8 Document procedures for data collection and handling

Much ancillary information is stored in the heads of experienced researchers which needs to be passed on in a structured and coherent way to the next generation. Procedures for data collection whilst at sea should be written up and made easily available, including check lists of information required to accompany data and samples of log sheets to be used. Instructions for processing procedures used in the institute should also be available. POGO institutes could develop a joint set of guidelines, based on those which exist at individual institutes (e.g. CSIRO, SOC). These should include instrument methods, standards and associated metadata.

9 The generation, management and preservation of metadata should be given high priority at the institutional level and POGO institutes should promote and work within broad (international) metadata models to ensure effective and efficient communication of data among institutes and internationally

The word “metadata” has more than one meaning. Firstly, there are many metadata standards (e.g. DIF, Dublin Core, ISO19115, FGDC, etc) which specify the pieces of information that need to be stored to describe a data set. However, the crucial point is that data collecting and holding organisations should maintain an inventory of their data holdings, including information about where and when the data were collected, what parameters were measured, a brief description of the data set and the physical location of data.

Secondly, metadata can refer to the supplementary information needed for a third party to use the data without reference to the data collector. This is particularly relevant for non-physical data. POGO institutes should ensure that all relevant information needed to accompany a data set is stored electronically along with the data set. This should include collection methods, protocols, data format description, processing methods, quality control methods, and unresolved problems with the data. The Data Type guidelines (below) include a check list of items to include for a variety of data types.

10 Use published guidelines/standards to process and quality control data

Where standards exist, these should be used to collect, process and quality control data. For example, both WOCE and JGOFS have published standards and protocols. Other schemes (e.g. QUASIMEME, BEQUALM, ISO standards, etc.) should be adhered to wherever possible. In addition, the ICES Working Group on Marine Data Management has published a series of Data Type Guidelines.

These Data Type Guidelines have been designed to describe the elements of data and metadata important to the ocean research community. These guidelines are targeted toward physical-chemical-biological data types collected on oceanographic research vessel cruises. Each guideline addresses the data and metadata requirements of a specific data type. The data types covered are: discrete water samples, CTD, XBT, net tow (plankton), moored current meter, shipboard ADCP, moored ADCP, surface underway measurements, SeaSoar/batfish, profiling float, surface drifting buoy and water level. These are available at: www.ices.dk/committe/occ/mdm/guidelines.

11 Ensure data are backed-up

Researchers within POGO institutes must ensure that their data are backed-up. For example, there are examples of students whose entire Ph.D. data set has been stored on a single floppy disk and many cases of data stored on PC hard disks which are not routinely backed-up. In addition, supplementary information, without which the data cannot be used, needs to be stored along with the data, and also backed-up.

12 Ensure effective data stewardship/archive and data access facilities

Data centres have expertise in stewardship and archiving, and their assistance should be sought by POGO institutes. This option should always be examined carefully, especially for long-term data stewardship. Data need to be transferred from one storage medium to the next, ensuring that data can be made available in an appropriate form in the future. Moreover back-up copies of the data need to be kept and disaster recovery systems put in place to ensure that data and accompanying information are not lost.

When data stewardship has been implemented, then consideration can be given to data access. Various options are available: data centres have a variety of data sets available on-line, OPeNDAP (formerly DODS) is used by a number of organisations and new systems are under consideration (e.g. US Integrated Ocean Observing System (IOOS), Data Management and Communications (DMAC)).

EDUCATION

13 Educate (new) staff/students

POGO institutes should try to ensure that researchers know what is required for good data management and why. For example, all too often students have sailed on cruises with little idea of what constitutes important metadata or why it should be recorded.

Mechanisms for education and training include holding a series of lectures/seminars as part of the academic curriculum for both undergraduate and post-graduate students, running hands-on training courses, and encouraging internet learning (e.g. using IOC's OceanTeacher [www.oceanteacher.org], or the Visual Geophysical Exploration Environment (VGGE)). In addition, POGO institutes can be twinned to spread the development of data management capacity and good practice. In addition, POGO institutes should have a key role in developing secondary and tertiary curricula and provide illustrative data for this purpose.

POGO institutes should also encourage the participation of their scientists in data management working groups (e.g. WOCE Data Products Committee, CLIVAR and Argo data management). They should also encourage participation in IOC/IODE groups (e.g. the Group of Experts on Biological and Chemical Data Management and Exchange Practices) and the ICES Working Group on Marine Data Management.
