

DATA AND SAMPLE MANAGEMENT POLICY FOR THE AMT CONSORTIUM

1. Introduction

1.1 NERC requires all thematic & consortia programmes to plan adequately for the management of the data they will collect. The planning must cover not only the practical arrangements while the programme is running, but also the subsequent maintenance of datasets of longterm importance. Properly managed, the data will provide a key NERC resource, which will continue to be used long after the formal end of the programme. The scale of effort dedicated to data stewardship should reflect the anticipated longterm value of the data.

1.2 In the context of NERC Data Policy, "data may be held in either analogue or digital form and be stored either on paper or a variety of computer-compatible media... physical specimens in curated collections are outside the usual sense of the word". Nevertheless, it is appropriate for sample management issues to be also discussed in this position paper, making clear which aspects apply in different sections.

2. Role of BODC

2.1 The British Oceanographic Data Centre (BODC) is the NERC Designated Data Centre for digital information arising from the AMT programme. BODC has provided data management services for many other multi-laboratory and multidisciplinary marine programmes, NERC and non-NERC (eg BOFS, LOIS, PRIME and OMEX), and has delivered high quality, accessible data sets, primarily via CDs, for further scientific use. Costs have been allocated in the AMT budget for BODC services, and Dr Castellani Claudia (cxc@bodc.ac.uk) and Dr Gwen Moncoiffe (gmon@bodc.ac.uk) have been appointed to collate and make accessible all AMT data (1995-2005).

3. Minimum standards of stewardship for NERC corporate data

3.1 The following minimum standards are expected to apply when (digital) datasets form part of NERC's enduring data resource:

- i) The ownership and Intellectual Property Rights to the dataset must be established, and NERC's policy towards exploiting and making it available to third parties agreed
- ii) The dataset must be catalogued to the level of detail required by a NERC Designated Data Centre, so that it can be mentioned in web-based NERC data catalogues
- iii) Formal responsibility for the custody of the dataset must be agreed
- iv) The data must be fully "worked up" (ie calibrated, quality-controlled etc) with sufficient associated documentation to be of use to third parties without reference to the original collector
- v) The technical details of how the data are to be stored, managed and accessed must be agreed and suitably documented
- vi) The technological implications must be established (digital data stewardship implies the need for an underlying infrastructure of IT equipment and support)

- vii) The resources needed to carry out these intentions over the planned life of the data, in terms of staff (Data Centre) and IT equipment/infrastructure must be estimated and sources identified.
- viii) A review mechanism must exist to reconsider periodically the costs and benefits of continuing to maintain the data. The intention to destroy or put at risk data should be publicised in advance, allowing time for response by interested parties.

3.2 The above NERC-wide requirements will be managed by BODC. Nevertheless, PIs need to be aware of this framework.

4. Data and sample acquisition

- 4.1 A well-structured and user-friendly identification system is essential for cruise-based data collection and sample labelling. Such arrangements are traditionally the responsibility of the cruise Principal Scientist. For AMT, an overall consistency in approach is necessary - with cruise identifiers linked to unique combinations of station and/or observation numbers. AMT protocols are being developed with direct BODC involvement.
- 4.2 BODC staff need to be directly involved in planning AMT fieldwork, eg with the opportunity to attend cruise planning meetings and Steering Committee meetings. BODC cruise participation is also welcomed (subject to berth availability), particularly when novel and/or relatively complex sampling arrangements are involved.
- 4.3 Station identifiers, navigational information and "basic" oceanographic data (for which BODC will have quality-control responsibilities) must be provided to BODC by the Principal Scientist immediately after an AMT cruise. A copy of the Cruise Summary Report (ROSCOP form) should be provided to BODC by the Principal Scientist within one week of the end of the cruise. A copy of the full cruise report should also be sent to BODC, preferably electronically, as soon as it is completed. BODC will then assist AMT in making this more widely available (eg via a link from the AMT website).
- 4.4 Processed and project-specific cruise data must be provided to BODC by the Principal Scientist and project teams within the timescales agreed and detailed below, not in the concluding few months or weeks of projects. However, great importance is given both by the programme and by BODC to protecting the interests of data originators, and restrictions on the wider availability of BODC-held datasets will therefore apply (see below).
- 4.5 BODC staff will ensure the effective linkage with the British Atmospheric Data Centre (BADC) to allow the atmospheric community access to the data.

5. Data formats and data media

- 5.1 Digital data should be collected and stored using standard, widely-available software products and their related data formats. Whilst BODC has experience in handling a very wide range of software, formats and media, PIs should discuss with them the proposed use of any data-handling or storage protocols that might be regarded as "non-standard".
- 5.2 CD-ROMs are currently the preferred means for making integrated data products from marine thematics available to the wider research community. The AMT Steering Committee will advise on the number of CDs, and set target times for their release.

6. Data back-up policy

- 6.1 Daily back-up programmes apply at BODC to safeguard major digital databases. Project PIs are responsible for providing appropriate back-up strategies for unique digital data stored locally and/or via other organisations.
- 6.2 As far as possible, analogue data (such as photographs) should be "disaster proofed" by transferring them into digital form, eg by scanning. Such duplication is not a waste of effort, even though the original, analogue version may have a longer lifetime than the format/media used for the digital transcription. Such data may then be included on a programme CD. Note that BODC has considerable experience in managing and publishing image data.

7. Protection of Data Originators' Intellectual Property Rights

- 7.1 The following arrangements have been developed to ensure an appropriate balance between the protection of data originators' intellectual property rights and the potential benefits that may arise via data use by the programme, the wider research community and other interested parties.
- i) All data collected in the AMT programme must be provided to BODC within the timescales agreed and detailed below, and will be freely available to all programme participants (PIs and CoIs) for AMT purposes on the condition that the originator is kept informed about how the data are being used and is duly acknowledged in any exploitation of that data
 - ii) Due acknowledgement is considered to be co-authorship, specific reference / acknowledgement to the data source or a share of any financial reward. The form of this should be negotiated between the data originator and the data exploiter. If a dispute should arise, then the problem will be referred to the Steering Committee for resolution.
 - iii) Data collected as part of an AMT studentship will be protected until submission of the thesis or until 4 years after the commencement of the PhD (whichever is sooner).
 - iv) Until AMT data enter the public domain, BODC will not transfer them to parties outside the programme without the explicit agreement of the originator. Steering Committee advice will also need to be sought if major data transfers are involved, to avoid compromising the interests of other programme participants.
 - v) The timescale for entry into the public domain is dependant upon parameter and is detailed in the table below. The mechanism for release of all AMT data to the public domain is expected to be the production of the AMT CD-ROM at the conclusion of the programme.
 - vi) A condition of CD-ROM usage is that it is regarded as a data publication and all usage of the data contained therein should acknowledge the data originator through citation

8. Longterm sample curation

- 8.1 Material collected on AMT cruises is owned collectively by the programme. However, during the programme lifetime, sample-originators have responsibility for the stewardship

of material, recording any removals and (if shared with other research groups) keeping track of its movements and usage.

- 8.2 It is recognised that indefinite storage of all biological material is impractical, and that some identification and analytic procedures require sample destruction. Nevertheless, it is expected that nearly all net-collected zooplankton and representative sub-samples of microplankton will be stored for the duration of the programme by sample-originators. Subsequently, longterm archiving (of at least 5-10 yr) will be arranged by the programme for as many samples as possible, to maximise the exploitation of the taxonomic information that they contain.
- 8.3 Before project PIs dispose of biological material in their possession, an assessment should therefore be made as to whether it might be of value to other research groups, not necessarily part of the AMT programme.

9. Data and sample availability

- 9.1 It is NERC policy to ensure that "individual scientists, principal investigator teams and participants in programmes will be permitted a reasonable period to work exclusively on, and publish the results of, the data collected by such individuals and teams". Nevertheless, as the AMT programme develops, there is necessarily a sequential widening of access to data and samples. This process has already been outlined with respect to data under 7 above. It can be generalised with reference to three access levels:

Level 1 (AMT Project). Availability limited to the investigators responsible for data/sample collection, any wider sharing at the discretion of the investigators

Level 2 (AMT Programme). When data is transferred to BODC, their availability is automatically extended to other investigators within the AMT programme. Nevertheless, their further use is still under the control of the data originator, and any wider sharing is at the discretion of the AMT Steering Committee.

Level 3 (Public). Data publication (depending on parameter ranging from real time to at or near the end of the programme). Availability extended to external users, either openly (for academic use) or at the discretion of BODC/NERC (for commercial exploitation, in consultation with data-originators). Post-programme availability of biological material to be controlled by the body responsible for its archiving, on the basis that 'ground-rules' will then have been established by the Steering Committee, and that sample-originators will be consulted wherever practicable.

- 9.2 It is to the benefit of the programme as a whole that the transition between Levels 1 and 2 is made as rapidly as possible.

10. Identifying data and samples for management purposes

- 10.1 It is important that the AMT programme maintains an awareness of all data and samples collected through its support, including outputs from partnership arrangements. Thus a requirement of participating in AMT cruises or having samples collected by AMT participants will be a report detailing such data/samples, and their stewardship arrangements, if not via BODC.

Table of AMT determinands and timescales of release to the AMT community and into the public domain e.g. real time (RT), 3 months after collection (3 mo), 1 yr after collection (1 yr) and 6 mo before the end of the programme (April 2006).

Determinand	Level 1 (Project / PI responsible)	Level 2 (AMT)	Level 3 (Public)
<i>Underway</i>			
T, S	BAS / BODC	3 mo	3 mo
Chl (fluor)	PH	3 mo	3 mo
CH ₄ , N ₂ O	CSL	6 mo	2006
pCO ₂	AJW / NL	3 mo	2006
FRRF	JA / SL	6 mo	6 mo
Lugols & formalin	PH	1 yr	2006
HPLC	PH	6 mo	6 mo
Bacterial numbers	MVZ	6 mo	2006
POC/N	RS		
MVP	EMSW/UKORS	3 mo	3 mo
Optics from MVP	JA / SL	6 mo	6 mo
Alkalinity (2 samples /day)	AJW / NL	3 mo	2006
DIC (2 samples /day)	AJW / NL	3 mo	2006
<i>Discrete</i>			
CTD T, S, transmission, PAR	BAS / BODC	1 mo	3 mo
CTD O ₂ sensor	CR / BODC	3 mo	3 mo
CTD Chl	PH / BODC		
Winkler O ₂	CR	3 mo	3 mo
Alkalinity	AJW / NL	3 mo	2006
HPLC pigments	PH		
Lugols & formalin	PH	1 yr	2006
Bacterial numbers	MVZ	6 mo	2006
POC/N	RS	6 mo	6 mo
¹⁵ N – PON & DON	RW	6 mo	2006
PIC	PH	6 mo	6 mo
DOC/N	RS	1 yr	1 yr
Micromolar nitrate, nitrite, phosphate & silicate autoanalyser	EMSW	1 mo	3 mo
Nanomolar nitrate, nitrite, phosphate & ammonium	EMSW	3 mo	3 mo
Nitrate +? sensor	EMSW	3 mo	3 mo
Size fractionated PP	PH	3 mo	3 mo
Size fractionated chl (fluorometry)	PH	3 mo	3 mo
DOC production	PH	3 mo	3 mo
Calcification	PH	3 mo	3 mo
Gross production, net community production & respiration (O ₂)	CR	3 mo	2006

Photodegradation of CDOM (O ₂ & DIC)	CR (1 mo)	3 mo	2006
CDOM absorbance	RU-G (1 mo)	3 mo	2006
DIC	AJW / NL (1 mo)	3 mo	2006
Optics (Lu, Ed & Es)	JA / SL	6 mo	6 mo
FRRF	JA / SL	6 mo	6 mo
N fixation	CSL / APRE	6 mo	2006
¹⁵ N nitrate, ammonium and urea uptake	ML / APRE	6 mo	2006
Bacterial biodiversity	MVZ	1 yr	2006
Bacterial activity	MVZ	1 yr	2006
Zooplankton biomass	XI	1 yr	2006
Microzooplankton biomass	EF/PHB/DM	1 yr	2006
Microzooplankton activity	PHB/DM/EF	1 yr	2006
DMS	GM	3 mo	2006
NH ₃	GM	3 mo	2006
Aerosol metal, isotope and major ion	TJ	1 yr	2006
Rainwater metal, isotope and major ion	TJ	1 yr	2006
Atmospheric / dissolved N ₂ O	RU-G	1 yr	2006
Stable isotope ratios of N ₂ O	RU-G	1 yr	2006
RS products	JA / SL	3 mo	3 mo (must be NASA authorised user for SeaWiFS)
