
# *RAPID Climate Change Programme*

Data Management Plan

## 1. Introduction

NERC requires all Thematic Programmes to plan and implement a data management scheme. The planning must cover the practical arrangements while the programme is running and the subsequent maintenance and long-term curation of the data sets. The latter is increasingly important in view of the Environmental Information Regulations, which place a duty on Government funded bodies to make all publicly funded data readily and easily available.

The NERC Data Policy requires that all data are lodged with the appropriate NERC Designated Data Centre. In the context of RAPID these are the British Oceanographic Data Centre (BODC) and the British Atmospheric Data Centre (BADC), the respective Designated Data Centres for Marine and Atmospheric Sciences. The minimum required standards of stewardship are summarised in section 3.

NERC provides funding to the Data Centres for basic infra-structure support and the long-term maintenance and curation of NERC's data assets. Thematic Programme budgets include the funds necessary for within project data management for the life of the project.

An integral part of the Data Plan is an obligation upon RAPID Programme Principal Investigators (PIs) to ensure that data management is undertaken in a suitable way, and that adequate consideration is given to the "data side" of their work. Individual project 'data management plans' will cover staff responsibilities, data collection policies, data standards, resourcing of data management, data quality and quality assurance.

The data management policy as defined by the RAPID steering group is outlined in Annex I.

This plan has been formulated following a review of the specified resource requirements and outputs set out in the project proposals and a series of discussions between BODC/BADC, the RAPID Science Co-ordinator (Dr Meric Srokosz) and the project PIs in order to assess the scale of data collection/production. These include observational and modelling products, physical samples and the requirements to enable links to third party data sets.

## 2. The Role of the RAPID Data Centre (RDC)

Submission of and access to data will be through a common 'portal' and for the purposes of RAPID the term RAPID Data Centre (RDC) will refer to BADC and BODC. Data management costs have been allocated in the RAPID budget for RDC services.

Given the complex and broad range of data encompassed by the RAPID programme the nature of the data management will vary between projects. The basis of this has been agreed with the PIs following an initial dialogue with the RDC and RAPID Science Co-ordinator. This is outlined in Annex III.

The RDC will be the focal point for PIs regarding data issues. The RDC website will contain inventories providing comprehensive up to date information about the status of all project data sets and model runs, so that all RAPID participants can easily request available data. The RDC will service data requests by RAPID participants.

Following the completion of RAPID the RDC will ensure that data are passed to the appropriate International Data Centres, ensuring that NERC meets its international obligations.

## 3. Minimum Standards of Stewardship for NERC Data

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

i. The ownership and Intellectual Property Rights to the data set must be established, and NERC's policy towards exploiting and making it available to third parties agreed.
ii. The data set 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 data set must be agreed.
iv. The data must be fully "worked up" (i.e. 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 (whether in project teams or the 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.

The above NERC-wide requirements, set out in the NERC Data Policy (www.nerc.ac.uk/data/policy.shtml), will be looked after "automatically" for the RAPID data sets managed by BODC and BADC. Nevertheless, PIs need to be aware of this framework.

## 4. Data and Sample Acquisition

RAPID data cover a broad subject area, including oceanographic data, sediment cores, palaeo data and the generation of model output. It is not the intention of this document to specify in detail how these data be collected, described and delivered to the data centres, however, a number of generic principles need to be adhered to.

Processed and project-specific data must be provided to the RDC by the Principal Scientist and project teams as they become available, not in the concluding few months or weeks of projects. However, great importance is given, both by the programme and by the RDC, to protecting the interests of data originators, and

restrictions on the wider availability of the RDC-held data sets will therefore apply (see the Data Policy at Annex I).

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. However, in order to assist the PIs and RDC it is necessary that a representative of the RDC attend the pre-cruise meeting.

Station identifiers, navigational information and "basic" oceanographic data (for which the RDC will have quality-control responsibilities) must be provided to the RDC by the Principal Scientist immediately after a cruise. Normal practice, as for other Thematic Programmes, will be for the RDC to meet the ship when it docks in the UK and to take delivery of this material together with a copy of the logs, calibration data and sensor information. If a cruise terminates in a foreign port it will be necessary for the PI and a representative of the RDC to meet immediately on return of the PI to the UK. A copy of the Cruise Summary Report (ROSCOP form) should be provided to the RDC by the Principal Scientist within one working week of the end of the cruise. A copy of the full cruise report should also be sent to the RDC, electronically, as soon as it is completed. The RDC will then assist in making this more widely available (e.g. via a link from the main programme website).

In the case of palaeo digital data, a representative from the RDC needs to be involved prior to and immediately after field campaigns in order to obtain the necessary information to describe the nature of the collected data. Where appropriate, both the raw and derived data will be stored together. For example, data resulting from analyses on an ice core would have raw parameters, such as depth in core, which would be stored alongside the derived parameters, such as age. Accompanying the data will be a description of the method used to arrive at the derived parameters.

For projects collecting physical samples it is the responsibility of the PIs to ensure that appropriate management measures are in place. However, it is important that the necessary collection details are provided to the RDC in order that the information forms part of the overall project information. For deep sea sediment cores, the samples must be deposited with BOSCOR (British Ocean Sediment Core Repository), hosted by the Southampton Oceanography Centre. Those dealing with palaeo samples (e.g. speleothem, bog cores and lake samples) will continue with existing management practices.

In the case of model data, the details for submission and serving will be agreed with individual PIs. Broad principles are given in section 5. In general, information accompanying submitted model data should include the model name and version number and a brief description of the model’s general aim. See the metadata protocol (Annex II) for more detail.

**Metadata**

Metadata are a crucial part of any data archive since they ensure that the data can be understood at a later date. To guarantee the RAPID data archive quality, full documentation on all validated raw and processed data, as well as on models and model results, must be provided to the RDC. It is therefore essential that metadata are

submitted at the same time as the data sets to which they pertain. The responsibility for producing the metadata will lie with project PIs and the RDC. A metadata protocol is outlined at Annex II.

In addition to the standard metadata, investigators are encouraged to archive at the RDC all relevant information electronically, including references, papers, reports, etc., unless agreed otherwise between the PIs and the RDC.

## 5. Data Formats and Data Media

Digital data should be collected and stored using standard, widely available software products and their related data formats. Whilst the RDC has experience in handling a very wide range of software, formats and media, Investigators should discuss with them at an early stage the proposed use of any data-handling or storage protocols that might be regarded as "non-standard".

In general, model data should be formatted in CF compliant NetCDF files, although there will be exceptions (particularly PP and HDF will also be accepted). Documentation on formats and conventions is available from the RDC (http://www.badc.rl.ac.uk/formats/), which also provides links to downloadable free software packages to support NetCDF access.

Submission of data will generally be via CD-ROM, as a Word/Excel e-mail attachment or by *ftp*. In some instances (e.g. some of the atmospheric model output) an automatic Web based file uploader will be available. At an early stage Investigators should discuss the options with the RDC.

CD-ROMs and or DVDs are currently the preferred means for making integrated data products from thematics available to the wider research community. However, there may be a preference towards a web-based final data product as RAPID progresses. The RAPID Steering Committee will review and decide on this at a later stage in the programme. It is not expected that the choice made will have cost implications.

## 6. Data Back-up Policy

The consequences of losing data, due to having made insufficient or inappropriate provision for their back-up, are potentially catastrophic in the case of large data collections, and cumulatively serious in the case of smaller data sets. Rigid daily back-up programmes operated at the RDC safeguard major digital databases. Provision and support of back-up strategies for digital data stored locally is the responsibility of individual PIs, or their delegates. Project PIs and Co-Is are responsible for providing appropriate back-up strategies for digital data stored locally and/or via other organisations.

As far as possible, analogue data (such as photographs) should be "disaster proofed" by transferring them into digital form, e.g. 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-ROM or DVD. Note that BODC has considerable experience in managing and publishing image data.

PIs should bear in mind that the timely deposit of data with the RDC will provide additional security for the project data.

## 7. Protection of Data Originator's Intellectual Property Rights (IPR)

The Steering Group and the RDC recognise the need to ensure reasonable protection of project scientist IPR. The RAPID Data Policy (see Annex I) addresses this and is intended to provide an appropriate balance between the protection of data originators' IPR and the potential benefits that may arise via data use by the programme, the wider research community and other interested parties.

## ANNEX I

## Rapid Climate Change (RAPID) Data Management Policy

This document describes the data management policy for the RAPID programme as drawn by the Steering Committee. The primary aim of the RAPID data policy is

- To encourage rapid dissemination of scientific results.
- To protect the rights of the individual scientists.
- To have all the involved researchers treated equitably.
- To ensure the quality of the data in the RAPID data archive.

These aims conflict at times, and it is hoped that the provisions of the protocol resolve these conflicts fairly. It is recognised that this cannot always be achieved to everyone's complete satisfaction; there are bound to be cases where individual interests clash with those of the RAPID programme. Therefore to try to meet these aims, all PIs involved in RAPID, in accordance with and on behalf of their co-investigators, have agreed to abide by the following conditions as part of the acceptance of the grant award:

### Data management

Data collected within the RAPID programme will comply with NERC's policy on data management (www.nerc.ac.uk/data/policy.shtml). The main objective of this policy is to ensure that the data will contribute to a key NERC resource, which will continue to be exploited both scientifically and commercially long after the formal end of the programme. The management of the data collected within the RAPID programme will be the responsibility of the relevant NERC Designated Data Centres (e.g. BADC, BODC), and funds have been made available from the RAPID budget to support this activity. In the absence of a NERC Designated Data Centre for palaeo data, special provision will be made for such data acquired within the programme (while ensuring appropriate links to international projects, such as HOLIVAR). To ensure proper data management the Science Coordinator will work together with a small data sub-group appointed by the Steering Committee (see appendix below).

### Recommended RAPID data policy (in line with other Thematics)

The data subgroup proposed the following data policy for the RAPID programme, which has been ratified by the full Steering Committee and will apply to all projects funded through RAPID:

a) Data[1] should be lodged with the appropriate data centre on acquisition[2], together with such metadata as are defined under the RAPID data management plan.

[1] **Data:** includes palaeo data, present-day observations, model output, data syntheses, data-model syntheses, model codes and physical samples.

[2] **On acquisition:** the time-scale may vary between data types (for example, real-time data could go directly to a data centre) but the overall aim is to keep the time-scale as short as possible and certainly less than 6 months. This is to ensure that data acquired during RAPID are available to the RAPID community within the lifetime of the programme.

b) Data will be embargoed for 1 year from acquisition, allowing the PI and co-workers to exploit it in the first instance. The metadata will not be embargoed, to allow the wider community to be aware of work being carried out under RAPID and facilitate community building.
c) Data will be made available to the RAPID community after 1 year, and to everyone after 2 years.
d) Anyone making use of RAPID data within 3 years of it being lodged at the data centre will be required to include the PI and/or co-workers (as appropriate) as co-author/s on any resulting papers, if the PI and/or co-workers so desire.
e) Any corrections, improvements or amendments to data must be lodged with the appropriate data centre as soon as possible.
f) PIs making use of RAPID data are responsible for ensuring that the data used in publications are the best available at the time.
g) Data submitted to the data centre must be in the data format agreed between the data centre and PI. In addition, all agreed metadata must be supplied to the data centre.
h) While data are restricted from the public domain, no data will be transferred to parties outside the programme without the explicit agreement of the originator. In addition, guidance will need to be sought from the Science Coordinator and the Steering Committee if major data transfers are involved, to avoid compromising the interests of other programme participants.
i) In the event of dispute, the final decision rests with the RAPID Science Coordinator and the Steering Committee.
j) PIs and/or co-workers failing to comply with the RAPID data policy would be subject to appropriate sanctions.

**Appendix: Data subgroup membership**

K. Briffa, P. Challenor, S. Tett, M. Srokosz, C. Gommenginger and a RDC representative.

# ANNEX II

# RAPID Metadata Protocol[1]

## 1. Introduction

The term *metadata* encompasses all the information necessary to interpret, understand and use a given dataset. *Discovery metadata* more particularly apply to information (keywords) that can be used to identify and locate the data that meet the user's requirements (*via* a Web browser, a Web based catalogue, etc). *Detailed metadata* include the additional information necessary for a user to work with the data without reference back to the data provider. The metadata required by the RAPID Programme include both discovery and detailed metadata.

Metadata pertaining to observational data, for example, include details about **how** (with which instrument or technique), **when** and **where** the data have been collected, by **whom** (including affiliation and contact address or telephone number) and in the framework of which research project.

In the case of all submitted data, the RDC needs to know how the values were arrived at. The derivation process must be stated: all processing and calibration steps should be described and calibration values supplied. The nature and units of the recorded variables are essential, as well as the grid or the reference system. The RDC requests that as much information as possible about fieldwork instrumentation be included, e.g. serial number, copies of manufacturer's calibration sheets, and recent calibrations, if applicable.

Metadata pertaining to model output should include the name of the model, the conditions of the calculation, the nature of its output, the geographical domain over which the output is defined (when applicable). Specific conditions applying to the model or the experiment may be mentioned. Metadata also include information on the format in which the data are stored, and the order of the variables, to allow potential users to read them. Metadata pertaining to software models include the key points of the theory on which the model is based, the techniques and computational language used, and references.

The following lists the minimum metadata required to accompany data files submitted to the RAPID Data Centre (RDC). Since there is a large range of data types within RAPID, the RDC will liaise with project workers submitting data on a case-by-case basis to ensure that metadata formats are appropriate and to gain additional relevant information as necessary.

## 2. Metadata for RAPID Projects

### 2.1 Metadata for tables of numbers (observations or model output)

#### 2.1.1 Content

Metadata include the following overall information. Some information in this list may be applicable in specific cases only.

[1] Adapted from URGENT Air Metadata document at http://badc.nerc.ac.uk/data/urgent/Metadata.html

- **Information about the experiment.**
  Date when fieldwork, experiment or model simulation started.
  Site or trajectory bounding box or domain limits.
  Platform (e.g. ship, cruise number), instrumentation (including instrument make, model and serial number).
  Model name.

- **Information about the experimenter(s).**
  Names, affiliation, contact address including e-mail, telephone number.
  Programme name, research project number.

- **Information about the independent variables (spatio-temporal grid).**
  Names, units, domain of definition of independent variables.
  Interval values when appropriate.

- **Information about the data, including processing level.**
  Version number.
  Date of last revision.
  Processing level (nature of raw data, derivation method: processing steps, calibrations applied).
  Nature, name, units, scaling factors of dependent variables.

- **Information about data storage.**
  Number of files of the entire dataset.
  File number of current file.

- **Information about data format.**
  Type of format e.g. ASCII, Excel, Matlab, NetCDF.

- **Additional information.**
  May include particular conditions of experiment or model run, model boundary conditions, article reference, and sources of further information.

### 2.1.2 Metadata storage

Ideally, each data file should include a header containing the metadata. If there is a large amount of information (e.g. description of many processing steps, calibration techniques), then a separate text file can be used as an alternative.

## 2.2 Metadata for software

### 2.2.1 Content

Metadata pertaining to a model should include the following.

- **Information on the model**
  Brief description of model general aim.
  Model structure.
  Physical processes involved, including equation set.
  Algorithmic implementation techniques used.
  Spatio-temporal coverage when applying.
  Boundary conditions, including reference(s).
  Initial conditions, including reference(s).
  Program language.

Input nature and format.
Output nature and format.
Summary of model validation, or appropriate reference(s).
Summary of results from former studies conducted with the model, or appropriate reference(s).

- **Information on the author(s)**
  Names, affiliation, contact address including e-mail, telephone number.
  Programme name, research project number.

### 2.2.2 Metadata storage

Metadata relative to software can be included as comments in the top section of the source file or can alternatively be provided as a separate text file.

### 2.2.3 Format

Text. There is no particular requirement regarding software metadata formatting.

## 3. Additional documentation

Any additional documentation on recorded data or images, whether pertaining to a single data file or a whole dataset, that would not find its place into the structures described above (because it does not fall into any described category or because it is too voluminous) may be submitted to the RDC in the form of a text file that will be stored in the RAPID archive documentation directory. These documents may for example include technique description, possible use of the data, study conclusions, etc.

# ANNEX III

## Summary of work by RAPID Data Centre for each project

## Introduction

This document outlines the work required of the RAPID Data Centre (RDC) for each project funded under the first AO, with *approximate* milestones. The contents are based on the original proposals submitted to NERC and subsequent discussions with the project PIs. **Please note**: The ability of the RDC to meet the milestones is entirely dependent on the co-operation of the individual participants.

Please refer to the Data Management Plan for an outline of the roles and responsibilities of the RDC and project participants in respect of data.

## NER/T/S/2002/425 – Andrew Willmott

**Project data summary**

The project involves running a regional model for the Barents Sea to examine and quantify the amount of dense water formed in polynyas and identify the pathways it follows. Sea ice processes will be parameterised at Keele and coupled with the POLCOMS model.

The runs will result in gridded daily fields of ice concentration and depth along with 3-D hydrography (salinity, temperature and flow fields) over approximately 34 vertical levels. Tidal constituents will be supplied for each point. No biology is involved (i.e. ERSEM). The meteorological forcing data and boundary conditions obtained from the NRL model (typically a few percent of the total output), will also be stored. The model runs will be typically a few years in duration, using scenarios of high NAO, low NAO and perhaps an 'average' year.

The regional model will test polynya parameterisation. The principal output from the work will be an estimate of the quantity of dense water formed in polynya systems over a freezing season. The model resolution covering the Barents Sea will be approximately 2-3 km, with perhaps a number of nested regions at 1 km to test sea ice parameterisations.

Data required include heat flux and wind stress estimates from ECMWF. Satellite SMMI data will be used for sea ice validation and will be sourced by the project. BODC will search for hydrographic data for the Barents Sea (CTD, current meters, XBT, tide gauge). Possible sources are VEINS, MAIA, the Russian data centre, and Levitus.

The main output will be approximately 3 or 4 model runs that will result in publications, and will be included in the RAPID dataset and final product.

**RDC tasks**

**Data sourcing**

- Search for and acquire Barents Sea hydrographic data (CTD, current meter, XBT, tide gauge).
- Assemble them into data sets, describe and store.
- Supply ECMWF data.

**Model output**

- Acquire model runs as NetCDF files, assemble metadata and advertise availability.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2004/05 | Complete data sourcing tasks |
| 2006/07 | Complete model output tasks |

## NER/T/S/2002/427 – Simon Josey

**Project data summary**

The project will focus on the determination of which patterns of surface and near-surface variables are related to (or are susceptible to cause) changes in the THC. The variables analysed will be the sea-air heat exchanges (long wave, short wave, and sensible heat and latent heat fluxes) and the fresh water exchanges (evaporation, precipitation). Patterns will be tracked in both observational data sets and model runs. Various experiments will be run with the coupled models (representing 10 000 years in total) in order to isolate modes of variation. The ocean component will then be forced by these modes (10 to 20 runs). Output will be kept at a monthly resolution.

Atmospheric and oceanographic data will be required for model validation. Resulting output will include frozen versions of models used plus model input.

Data that will be produced include:

- 1000-year control run of FORTE at a monthly time resolution (160 Gbytes). Can (should) be archived quickly.
- Regarding other model output, some investigation on the best archival approach is needed (archival of code and code input versus archival of output at reduced resolution, e.g. yearly rather than monthly, with more detail over time periods of special interest). The approximate time scale for archival is 18 months after project start.
- Analysis of existing observed surface fluxes from the International Comprehensive Ocean-Atmosphere Data Set (I-COADS)

**RDC tasks**

**Data sourcing**

- I-COADS data available from I-COADS web site.
- NCEP reanalysis (monthly updates), available from NCAR: surface forcing fields.
- Possibly ERA-40 (meteorological variables, air-sea exchanges), available from the BADC.
- Temperature and salinity data from the BODC.
- 1000-year HadCM3 output (currently restricted to COAPEC participants).
- 1000-year HadGEM run very desirable when model will be operational.

**Model output**

- Acquire model runs, assemble metadata and advertise availability.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2004/05 | Complete 1st five data sourcing tasks |
| 2005/06 | Provision of HadGEM 1000 yr run output, if available |
| 2006/07 | Complete model output tasks |

## NER/T/S/2002/430 – Eric Wolff

**Project data summary**

The project involves subsampling the GRIP and NGRIP ice cores, held in Copenhagen. The subsamples will be analysed at high resolution in transition periods for the Dansgaard-Oeschger event 8 (DO8) and the 8.2 kyr event. Although the ice cores were collected under European projects, the analyses carried out here will be regarded as RAPID data. The nature of the analyses means that the samples are destroyed in the process and no physical samples will remain.

There will be reappraisal of existing lower resolution data, much of which are publicly available at the Boulder repository, although there are European data that may not yet have been lodged there. However, the project already has access to these. Other palaeo data from RAPID projects will be made available through the Data Centre.

The observations will be compared with freshwater forcing and isotope-enabled HadCM3 model runs generated by Paul Valdes, which will be made available for use by the RAPID community. Time resolution will be monthly at both ends of the events, less elsewhere. Additionally, there some BAS model experiments will take place, though the details are not yet clear.

The volume of data generated by the project will be relatively small, of the order of a few thousand values for approximately 6-7 analytes. The data are likely to be submitted in two stages: the 8.2 kyr data will be submitted within a matter of months, followed by the DO8 data at a later stage. Low-level metadata will be made available as samples are obtained.

**RDC tasks**

**Palaeo data**

- Acquire 8.2 ky and Dansgaard-Oeschger spreadsheets with ice core data, measured and derived parameters, and associated detailed metadata describing the process by which the values were arrived at.
- Load data into database and advertise availability.
- Prepare detailed data document.

**Model output**

- Acquire BAS model runs, assemble metadata and advertise availability.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2004/05 | Bank 8.2 ky data set |
| 2005/06 | Bank Dansgaard-Oeschger data set |
| | Complete model output tasks |

## NER/T/S/2002/436 – Nick McCave

**Project data summary**

Fieldwork-generated data will result from box and Kasten cores at approximately 20 sites, though not all cores are likely to be suitable for full analysis in the context of RAPID. At these sites there will be CTD casts for temperature, salinity, and optical properties (transmissometer and nephelometer). Calibration of the CTD and ship's underway data will be undertaken by BODC. Coring sites will be selected with the aid of swath bathymetry if Charles Darwin is available, otherwise, 3.5 kHz echo sounder data will be used. If swath bathymetry is available, representative regions will be fully worked up by UKORS, whilst the remainder will be submitted to BODC with the necessary metadata. Similar data were collected on the Norwegian cruise in July 2003, where 6 multicores and 2 gravity cores were obtained. These data will also be regarded as part of the RAPID data set.

Existing data required for the project include historic current meter time series that have been identified in the expected coring areas. The criteria for usable records are that the time series exceed one year and that the meters were within 100 m of the seabed. BODC will source the data where necessary and seek further data in the specified regions. Additionally, the current meter data from the Norwegian continental shelf/slope pertinent to the July 2003 cruise should be considered for inclusion within RAPID.

**RDC tasks**

**Ship data**

- Acquire one cruise of unprocessed CTD, water bottle sample analyses, and underway.
- Acquire swath bathymetry data as processed x, y, z co-ordinate data and create metadata descriptor. Archive remaining raw data as accession
- Acquire bed-hop camera images, if taken, and create metadata descriptor.
- Load water bottle data into database.
- Transfer CTD, underway, to BODC's format.
- Visually screen CTD, underway and flag any data that appear to be non-oceanographic.
- Load CTD and underway data to the project database.
- Determine CTD and underway calibrations from water bottle samples and apply.
- Prepare detailed documentation (metadata) describing the instrumentation, deployment, processing, calibrations, data quality and problem areas.
- Advertise availability of dataset.

**Core data**

- Acquire fully processed data from cores (~10-20), with documentation detailing the collection, processing and calibration steps.
- Load the core data into the database and advertise availability.
- Prepare metadata documentation.

**Third Party data**

- Acquire Norwegian cruise CTD x 6, underway and current meter data from July 2003 cruise coring sites.
- Transfer all series to BODC format.
- Visually screen the data for spikes of non-oceanographic origin and apply flags.
- Prepare detailed documentation about the sites, moorings, instruments, processing, calibrations, quality control procedures and data quality.
- NODB banking.
- Search for and acquire existing current meter time series at expected coring areas.
- Create metadata descriptor for data sets.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2004/05 | Complete third party data tasks |
| | Bank 1 UK cruise |
| 2007/08 | Bank core data |

## NER/T/S/2002/439 – Ric Williams

**Project data summary**

Two modelling studies will be carried out: an idealised process study and a palaeo-circulation experiment. The process study involves running an isopycnal model (1.4° resolution and 6 vertical layers) for the North Atlantic to examine the processes behind observed signals in monitoring arrays. It is not yet clear what the suite of runs will comprise, however, the timescales of interest are season to decade. The palaeo-circulation study will be carried out within a tied studentship, and will examine glacial-interglacial changes in circulation and the effects on carbon draw down. The project starts in January 2004, lasting for 3 years, but the tied studentship will not begin until October 2004. It is likely that the project will be extended to account for the mismatch in start dates.

Three key requirements were noted for the model setup:

- Realistic topography - BODC will supply the latest GEBCO CD-ROM.
- ECMWF reanalysis.
- Palaeo-atmospheric data - Paul Valdes has agreed to supply this, and it should be accessed centrally (i.e. via BADC).

The process study will interpret observations from the monitoring arrays, primarily the Hughes project, and also the Toole and Cunningham arrays. These will be accessible from BODC, who will acquire them as they become available. Also of interest were ASOF data, if available, for information on boundary conditions.

The main output is likely to be model runs from the tied studentship, and possibly a suite of runs from the process study model. It is expected that these will be included in the RAPID dataset and final product.

**RDC tasks**

**Data sourcing**

- ECMWF reanalysis data.
- Palaeo-atmospheric data from Paul Valdes.
- Observations from Hughes, Toole and Cunningham arrays.
- ASOF data from ICES 2003-2006.
    - Archive data as accession, treat as special dataset and create metadata descriptor.
- Watson tracer data if pilot study successful.

**Model output**

- Acquire model runs as NetCDF files, assemble metadata and advertise availability.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2004/05 | Complete data sourcing tasks |
| 2006/07 | Complete model output tasks |

## NER/T/S/2002/440 – Keith Briffa

**Project data summary**

Ocean-atmosphere general circulation models (OAGCM) will be used to create synthetic proxy data in order to quantify their ability to represent climate variations over the last 1000 years. Existing and new proxy data will be used to reconstruct climate variations at optimal temporal (seasonal) and spatial resolutions. Existing chronologies will be borrowed from past research using dating techniques.

**Models used**

- HadCM3, run at the Hadley Centre by Michael Vellinga and Richard Wood. Results will be provided by Simon Tett (Hadley Centre).
- ECHO-G (ECHAM4 + HOPE), developed at the MPI Hamburg and run at the GKSS. Results will be provided by Hans von Storch (GKSS) and Ulrich Cubasch (Berlin) through the EU SO&P project (started in November 2002). The possibility of archiving the ECHO-G output in a RAPID archive is not clear, and will in any case depend on the approval of the participants to the SO&P project and of the concerned scientists at MPI and GKSS (Tim Osborn will raise this issue with SO&P collaborators at the Berlin meeting).
- Possibly other OAGCM: if run during the project, collaboration will be sought with groups performing relevant runs.

**Model experiments**

- 1000-yr control runs (such runs have already been completed with both HadCM3 and ECHO-G).
- Simulations of the climate evolution under external forcings including natural and anthropogenic factors (orbital, solar irradiance, volcanic activity, greenhouse gas emissions, tropospheric aerosol, ozone, land use changes). These will include:
    - Already existing (or currently underway) simulations of the last 500 years (HadCM3) and last 1000 years (ECHO-G).
    - Simulations covering the time period 1750-2000 (with the 2 models).
    - Selected computed variables will be archived at a monthly time resolution.

**Observational data used**

Many records will be retrieved from the National Climatic Data Centre (NCDC) in the US (land data) and from the Network for Geological and Environmental Data (PANGAEA) in Germany (ocean data).

- Early historic instrumental data: gridded monthly temperature and precipitation at a high time resolution over the last 250 years.
- Surface climate proxy data (tree-ring records) partly already archived at CRU, partly provided through SOAP, based on data collected in the US and archived in the International Tree-Ring Data Bank at NCDC.
- Existing calibrated climate reconstructions (temperature, moisture) based on the latter.
- New tree-ring data will be made available during the early phase of the project, originating from Scandinavia, the Alps, Northwest Africa and Canada.
- Ice core data from the Greenland Ice Core Project (GRIP) and the Greenland Ice Sheet Project (GISP), provided by Sigfus Johnsen (+ new isotope GRIP data).
- Ice sheet data over Greenland collected under the Program for Arctic Regional Climate Assessment (PARCA), provided and analysed by Ellen Mosley-Thompson.
- Low resolution (annual or lower) data from land and lakes, including borehole records (rock cores), reconstructions derived from fossils, speleothem data, pollen records from the Global Pollen Data Base in Marseille.

- Sediment records from saline lakes in Spain, through collaboration with Steve Juggins, University of Newcastle.
- Data from the European Pollen Database in Arles.
- Data from the Multi-Proxy Database System (MPDS) for Mediterranean and African palaeoclimatologists (Medias-France).
- Any new proxy data generated by RAPID projects during the course of the programme.

**Newly generated data**
The records listed above will be integrated into a 5° x 5° proxy data base, which will be complemented by spatio-temporal climate reconstructions (temperature, humidity, precipitation, precipitation-derived variables, such as drought indices) derived from the proxies over large areas including the North Atlantic, North America, Europe, Asia and North Africa. Error bars will be estimated. Data will be lodged with the NERC designated data centre, with the World Data Centre for Palaeoclimatology hosted by the US National Climatic Data Center (NCDC) and with CRU, which will also distribute them.

**RDC tasks**
**Palaeo data**
- Archive derived palaeo dataset.
    - Acquire gridded product ~ 3 variables on a 5° grid in N. Atlantic.
    - Archive in originator's format.
    - Create detailed metadata description and advertise availability.

**Model output**
- Archive final model runs as NetCDF, assemble metadata and advertise availability.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2006/07 | Archive final palaeo dataset |
| | Complete model output tasks |

## NER/T/S/2002/441 – Brian Hoskins

**Project data summary**
The project will be looking at ERA-40 reanalysis data, which BADC are already committed to serving. There will be an Atlas produced from the ERA-40 data set, available in November 2003 through the European Centre.

The HadAM3 model will be used in two modes
1) Full climate.
2) Simplified geometry (aqua planet).

It is unlikely that there will be much interest in the highly idealised aqua planet output for the rest of the RAPID community. However, some of the palaeo projects may be interested in storm tracks and moisture pattern effects and access to the output from the full climate model runs would be useful.

It is possible that a standard set of atmospheric diagnostics will be applied to other models such as HadCM3, HadGEM and HiGEM. The HadCM3 1000 year dataset produced for COAPEC is archived at BADC, though there are gaps in the data. It is hoped that it will be made available to RAPID (negotiations are currently underway). The forthcoming 1000 year HadGEM run, being a Hadley Centre run, may need to be paid for and additional funds could be sought from RAPID. Since HiGEM will be run under a NERC consortium, these data should be freely available, although possibly not in time for this project.

It is not yet clear what model output should be included on the final product. This will emerge as time progresses.

**RDC tasks**

**Model output**

- Archive HadAM3model runs, assemble metadata and advertise availability.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2006/07 | Complete model output tasks |

## NER/T/S/2002/442 – Eric Guilyardi

**Project data summary**

This project will mainly use the HadCM3 1000 year run, which is archived at BADC through COAPEC. It is hoped that it will be made available to RAPID (negotiations are currently underway). Other UK community models such as HadGEM and HiGEM may also be used but their availability in time for this project is not yet certain. The PI also has access to models from allied research projects such as PRISM and CAPRI.

It is yet to be determined what detailed experiments will be undertaken, but broadly it is expected that these will comprise core experiments with a few long integrations (a few hundred years) and a number of shorter ones (several decades).

It is not yet clear what model output should be included on the final product. This will emerge as time progresses.

**RDC tasks**

**Model output**

- Archive model runs, assemble metadata and advertise availability.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2006/07 | Complete model output tasks |

## NER/T/S/2002/443 – Sandy Tudhope

**Project data summary**

**Note:** The RDC was unable to meet with PI Tudhope to discuss data management before preparation of this document. However, a brief opportunistic discussion with Co-I Collins was held in July 2003, and the modelling component of the project was described. The palaeo data estimates are summarised from the original proposal and are assumed to not greatly differ from what is currently planned.

The project will use a combination of palaeoclimate reconstruction from annually-banded corals and the fully coupled HadCM3 GCM to examine ENSO variability. Three periods are targeted: 0-2.5 ka; 6-9 ka; and 2000-2100 AD. The model experiments will result in one or two major integrations ~ 100 years long about 6000 years ago and 2-3000 years ago, depending on the palaeo data results. HadCM3 (1000 year) will be used with the standard set of diagnostics. Modelling work is expected to start in Autumn 2003. In terms of external data required for the modelling aspect, forcing will obtained from Paul Valdes, but no other data are required from elsewhere.

With respect to observational data, existing coral cores from a collaboration with an NSF-funded project (PI Pandolfi) and cores obtained in July 2002 will be dated and preliminary analyses carried out before a final fieldwork season targeting collection of material of specific ages. Optimal cores for further analyses will be selected using high precision U-series dating. Salinity and SST proxies will result from analyses involving $\delta^{18}O$, Sr/Ca, Mg/Ca and U/Ca. Proxy records external to the project will be compiled from published and new work on other coral sites, and deep sea sediments, tree ring, lake sediment and geoarchaeological data from relevant regions.

The RDC assumes that the final palaeo dataset will be relatively small (several columns and a few thousand rows) and will be submitted as spreadsheets, accompanied with detailed information describing how the values were arrived at.

The anticipated final model output is several HadCM3 100 year runs and some shorter sensitivity studies. Other runs will be determined by the palaeo-observational data. It is likely that other RAPID PIs will be interested in the results.

**RDC tasks**

**Palaeo data**

- Acquire spreadsheets with palaeo data, measured and derived parameters, and associated detailed metadata describing the process by which the values were arrived at.
- Load data into database and advertise availability.
- Prepare detailed data document.

**Model output**

- Acquire model runs and archive, possibly with frozen versions of code.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2006/07 | Bank palaeo data |
| | Complete model output tasks |

## NER/T/S/2002/446 – Andy Watson

**Project data summary**

The modelling project will involve existing data from a variety of sources, including TRACTOR, ESOP2, and ARCICE, for which the PI possesses the SF6 data. As ARCICE was NERC-funded, the data should now be available for wider use. Equally, that from ESOP2 should be available as this EU project has now ended. However, there may potentially be delays in obtaining the data from TRACTOR as it not scheduled to finish until the end of 2003 and there is also US participation. We understand that the project already has access to these datasets but, in the wider context of RAPID, it would be sensible for them to be collated and archived centrally by BODC.

There will be inverse, box model and high resolution model output. In terms of storage, the high resolution modelling will produce the greatest volume, whilst that from the inverse technique will be modest. Provisionally, the dimensions of the high resolution are ~200 $km^2$ at a horizontal resolution of 1 km, with approximately 40 levels in the vertical. Typically, there will be 6-10 parameters. At this stage, the intention is to run the model for two years. Most likely, one will be the -ve NAO year of 1996, for which the PI has data on convection and the other will be for a +ve NAO year. The boundary conditions will be supplied by the Norwegian Nansen Institute. The output (tracer distribution) from the inverse/adjunct methods is likely to be compared with OCCAM (or equivalent) and it was suggested that some form of visualisation of the resulting comparisons might also be stored. Model output is expected to be available in years 2 and 3 of the project.

**RDC tasks**

ARCICE data

- JR44 cruise – 3000 SF6, 1500 dO18 (not at BODC).
- CTD data 151 stations (soc020016) on BODC database.
- CEFAS water sampler ~ 50 samples.
- Acquire all water sample data.
- Load to database.
- Create documentation.

**Data sourcing**

- TRACTOR, ESOP2 from ICES

    - Investigate availability of and acquire TRACTOR (ends Jan 2004) and ESOP2 CTD and tracer data.
  - Archive as special dataset with supporting documentation.

**Model output**
- Archive model runs, assemble metadata and advertise availability.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2004/05 | Complete data sourcing tasks |
| | Complete ARCICE data tasks |
| 2005/06 | Complete model output tasks |

## NER/T/S/2002/448 – Ian Fairchild

**Project data summary**
Data output from ASCRIBE will result from analyses of speleothem from which palaeoclimatic proxies will be derived. While the output will be relatively small, in terms of volume, a key issue is that a large amount of information needs to be stored with a relatively small amount of data.

There is some foreign collaboration at the sample acquisition level. The analysis of these samples will take place mostly in the UK (under RAPID funds) and in Ireland through collaboration with Frank McDermott, who is externally-funded. However, the output will be viewed as one dataset, not as separate Irish and UK entities, and will be included in the final RAPID dataset.

**RDC tasks**
**Palaeo data**
- Acquire spreadsheets with speleothem data, measured and derived parameters, and associated detailed metadata describing the process by which the values were arrived at.
- Load data into database and advertise availability.
- Prepare detailed data document.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2005/06 | Bank palaeo data |

## NER/T/S/2002/449 – Harry Bryden

**Project data summary**
Historical data will be collated along the approximate line of latitude 26°N, in order to extend the Cunningham MOC measurements backwards in time. The data comprise CTDs, XBTs, current meter time series and altimetry.

Much of the data for the western side of the Atlantic will be obtained from Bill Johns, who has indicated to the PI that data were available for use, even if they were not yet lodged within a data centre. The eastern side data would comprise the Canary Islands CTD time series and mooring array data, and the subsequent ESTOC data. It is possible that elements of the CANIGO data set are relevant, which can be freely obtained from the project CD available from the Irish Marine Data Centre.

Data sets are presently being accumulated from the WOCE DVDs and Levitus and it is understood that all these data will be archived at BODC and form part of the RAPID data set.

Broadly, the quantities of data anticipated are:
- Four transatlantic sections of 100 stations of CTD/XBT data (1957, 1981, 1992 & 1998).

- Western Atlantic 10 years of data at 4 moorings, each with 3 current meters (120 years of records).
- Miami CTD data with 20 sections and approx. 15 stations (~300 CTDs).
- Eastern Atlantic time series with a CTD station every two weeks from 1980 to present (~500 CTDs).
- Eastern Atlantic 10 year mooring array with approximately 15 instruments (~150 years of records).
- Altimeter sea level anomaly data from the Rennell Division since 1992 to present, with approximately 60 points every 10 days.
- CANIGO data and any other available data within 24-26.5°N.

The collated datasets will be checked for uniformity of quality and will then be submitted to BODC for later dissemination to the RAPID community.

**RDC tasks**

- Acquire verified collated dataset: as many PIs may require access to these data.
- Archive as special dataset that will be served out in the submitted format, add metadata description.
- Acquire final dataset (with derived products)
- Archive as special dataset that will be served out in the submitted format, add metadata description.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2004/05 | Archive collated data |
| 2007/08 | Archive final dataset |

## NER/T/S/2002/450 – Peter Challenor

**Project data summary**

This project will use the C-Goldstein coupled ocean-atmosphere GCM and the higher resolution HadCM3. Ensemble runs will be computed for the energy-balanced case and for scenarios extending over time periods of 100-150 years (with gas emissions from Tyndall Centre).

The data and software to be archived include

- Emulator input (20-D space) and uncertainties.
- Gas emission scenarios (a given % per year increase for each gas considered).
- ??Frozen version of the GCMs (a few kbytes).
- Emulator code + utilities to run it.
- Emulator output: either global averages or global or regional 3-D fields (uncertain as yet).

**RDC tasks**

**Data sourcing**

- COAPEC 1000-year HadCM3 output
- 1000-year runs of HadCM3 on Beowulf
- Other atmospheric data held at BADC (to calibrate the models — time scale: June 2004)
- Hydrographic data from WOCE DVD, Harry Bryden

**Model output**

- Archive HadCM3 runs, frozen versions of C-Goldstein, gas emission scenarios, climate model emulator and climate model emulator output.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2004/05 | Complete data sourcing tasks |
| 2006/07 | Complete model output tasks |

## NER/T/S/2002/453 – Sheldon Bacon

**Project data summary**

Data collection activities will consist of 2 cruises: 1 hydrographic and 1 coring. The hydrographic cruise will involve approximately 80 LADCP and CTDs with water sampling, and approximately 1 month underway data including flow-through, meteorological, ship ADCP (SADCP) and possibly swath bathymetry.

A mooring array with 6 sites will be deployed in summer 2004 and potentially be recovered in 2005, however, it may be turned around and remain out for a further year. **Note:** There are no planned data collection activities on any mooring servicing cruises, however, BODC have assumed that CTD calibration casts will be undertaken on turnaround/recovery.

Coring work will be undertaken on a RAPID cruise led by Nick McCave (NER/T/S/2002/436). Several sediment cores will be obtained, the number of which will be determined during the cruise. The physical cores will be submitted to BOSCOR, and the CTD, LADCP, underway and ship ADCP data will be submitted to BODC once worked up and calibrated.

Third party datasets the project will use include an IFREMER mooring array (OVID/CLIVAR) with 5 mooring sites; the CEFAS array (ASOF) upstream; and potentially an offshore Dutch mooring array (LOCO).

**RDC tasks**

**Ship data**

- Acquire one cruise of fully processed and calibrated CTD/LADCP (~80), water bottle sample analyses, underway, SADCP (~30 days) with documentation detailing all the collection, processing and calibration steps.
- Acquire swath bathymetry data as processed x, y, z co-ordinate data for banking, and remaining raw data as accession.
- Load water bottle data into database.
- Transfer CTD, underway, SADCP from Pstar format and LADCP from MatLab format to BODC's format.
- Visually screen CTD, underway, LADCP and flag any data that appear to be non-oceanographic. SADCP data will be screened in less detail and only large spikes flagged.
- Load CTD, LADCP, SADCP and underway data to the project database.
- Check calibrations against bottle samples provided.
- Prepare detailed documentation (metadata) describing the instrumentation, deployment, processing, calibrations, data quality and problem areas.
- Advertise availability of dataset.

**Moored instrument time series**

- Acquire fully processed and calibrated moored instrument time series with documentation detailing all the collection, processing and calibration steps: 14-20 RCMs, 1 MMP that profiled 1 x per day over approximately 1000m. The array will be deployed for 1 year.
- Transfer all the series into BODC format.
- Visually screen the data for spikes of non-oceanographic origin and apply flags
- Prepare detailed documentation about the site, mooring, instruments, processing, calibrations, quality control procedures and data quality.
- NODB banking and advertise availability.

**Core data**

- Acquire fully processed data from cores obtained on a shared cruise with Nick McCave (NER/T/S/2002/436), with documentation detailing the collection, processing and calibration steps.
- Load the core data into the project database and advertise availability.
- Prepare detailed data documentation.

**Third Party data**

Search for and acquire third party datasets.

- IFREMER mooring array over 5 sites with approx 14 RCM, 14 SeaCat CTD, 2 ADCP.
- CEFAS ASOF upstream array until 2006.
- 8 MicroCat, 2 RCM8, 2 HOMER, 1 ADCP per year.
- Dutch LOCO array 2003-2008 – 3 profiling CTDs.
- Archive as special dataset.
- Create metadata descriptor for data sets.

**Major assumptions**

- Array will be recovered in 2005.
- Calibration casts will be done on turnaround and recovery, but no other data collection will take place.
- Third party data will be procured for period when mooring array was out, but will be held in originator's format, not banked.
- CEFAS and LOCO time series estimates are from CLIVAR web page and are assumed to be correct.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2004/05 | Bank 1 cruise |
| 2005/06 | Bank recovery cruise calibration CTDs |
| | Bank moored array time series |
| | Complete third party data (2004) tasks |
| | Bank core data |

**If the deployment is extended by another year, the above table will be augmented to include:**

| | |
|---|---|
| 2006/07 | Bank recovery cruise calibration CTDs |
| | Bank moored array time series |
| | Complete third party data (2005) tasks |

## NER/T/S/2002/455 – Neil Wells

**Project data summary**

Existing data will be collated, including ARGO, PALACE and SOLO float data, XBT data and possibly altimeter data, all between 24 and 60 degrees north, resulting in approximately 50,000 temperature and salinity profiles.

The ARGO data will be sourced from the relevant data centres, with advice from BODC as necessary. PALACE/SOLO data will be obtained from the WOCE DVD. XBT datasets of interest include the Bill Johns regular Miami-Gibraltar XBT transects.

Other data sought include "classical" data sections. The Bryden project will be pulling historical datasets together, which will ultimately be available as part of RAPID. The various datasets used will be checked to ensure they are homogeneous and consistent and therefore suitable for use as a continuous dataset, before submission to the RDC

The main output will be a gridded time series of heat storage and heat fluxes along East-West sections, at approximate intervals of 5°. This will be included in the final RAPID dataset, along with the data from which it was produced.

**RDC tasks**

- Acquire reformatted, homogeneous version of the data collated, i.e. the input version and also the output version, which includes the gridded time series of heat flux and storage.
- Archive as special dataset.
- Create metadata description and advertise availability.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|

| | |
|---|---|
| 2005/06 | Archive final dataset |

## NER/T/S/2002/459 – Chris Pain

**Project data summary**

This project involves a fine scale adaptive ocean model to parameterise deep convection and will generate a large, and presently undetermined, volume of data. The principal output from the project will be an improved description of deep convection leading to greater understanding and parameratisation in more conventional OCMs. The area of interest is the GIN and north Atlantic, but the boundaries are not yet determined.

The innovative nature of the project means that traditional methods of data delivery and visualisation may not be appropriate, and it is currently unclear what method will be used. It is also not yet clear what model output should be included on the RAPID final product. This will emerge as time progresses.

**RDC tasks**

- Acquire and archive model runs, model grid, animations and assemble metadata.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2006/07 | Archive model output, grid and animations |

## NER/T/S/2002/460 – Jonathan Holmes

**Project data summary**

The ISOMAP project will use both modelling and observational data. The latter comprises existing and new data to be collected during the project duration. Broadly, observational data will be obtained from palaeoarchives (speleothem, cores from bogs and lakes) and modern water monitoring (cave drips, bog and lake water).

In addition, there are data that have already been collected in other projects. These include Jim Marshall's NERC-funded Hawes Water data, speleothem data from ASCRIBE's Crag Cave site, a Belgian speleothem site and a southern German lake site (Ammersee). It is expected that these data will form part of the RAPID data set, but the situation with respect to the Belgian and German data needs to be formally clarified. **Note:** A recent communication from the PI suggests that the Belgian and German data will not be included in the final RAPID data set, but the derivations may be.

Observational data will be submitted to the Data Centre in the form of spreadsheets containing measured parameters and inferred values, as they become available, but access will be restricted by requiring direct permission of the project PI. The observational data from ISOMAP UK will relatively small, in terms of volume. However, a key issue is that a large amount of information needs to be stored with a relatively small amount of data.

The modelling component of ISOMAP involves adding an isotope module to a coupled ocean atmosphere model (HadCM3, FAMOUS) to produce a 3-D isotopic distribution. The primary focus within ISOMAP will be the 8.2 kyr and Younger Dryas events. It is anticipated that several RAPID projects will be interested in the output. Observational data from ISOMAP and ASCRIBE will be used to validate the model. Eric Wolff will also evaluate the model against isotope distributions from ice core data.

For the 8.2 kyr event, an ensemble of simulations (approximately 200 years) will be run using FAMOUS. The Younger Dryas event will require fewer but longer (~5000 years) FAMOUS simulations. If there is sufficient time, the past 1000 years will be simulated by running HadCM3 with the isotope module. The envisaged timescale for full scheme isotope-enabled model output is in the third year of the project.

Mechanisms for making the model output available to other projects was discussed but not finalised. It is likely that a version of the model will be archived along with some averages, transients and some example isotopic distributions.

**RDC tasks**

**Palaeo data**

- Acquire spreadsheets with monitoring time series for bog, lake and caves.
- Acquire spreadsheets with palaeo data including measured and derived parameters, and associated detailed metadata describing the process by which the values were arrived at.
- Load data into database and advertise availability.
- Prepare detailed data documents.

**Model output**

- Archive HadCM3 runs and possibly frozen versions of code.
- Archive FAMOUS runs and possibly code.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2005/06 | Bank monitoring data |
| 2006/07 | Bank palaeo data |
| | Archive model output |

## NER/T/S/2002/462 – Jonathan Bamber

**Project data summary**

This project consists of a series of model experiments, using HadCM3, FAMOUS and EMIC, which incorporate an ice model. It will result in several runs that are of interest to the RAPID community. Access to the output has not yet been finalised. The project Co-I, Paul Valdes, has web pages designed for registered users to view and access the output. However, in the interests of RAPID, it would probably be better to have these centrally available from the RDC.

**RDC tasks**

**Model output**

- Acquire HadCM3, FAMOUS and EMIC model runs, assemble metadata and advertise availability.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC Task |
|---|---|
| 2006/07 | Archive model output |

## NER/T/S/2002/00481 – Stuart Cunningham

**Project data summary**

Mooring arrays, covering western, eastern and mid Atlantic sections, will be deployed over a 4 year period. The moorings will be serviced and data downloaded annually. The first deployment will take place in January 2004 on RRS Discovery. The western array will be subsequently serviced by US ships, in collaboration with NSF-funded Bill Johns. The servicing of the mid Atlantic and eastern arrays will be carried out on RAPID cruises.

Immediately after the deployment cruise, a JRD cruise, approximately 37 days duration, will complete a trans-Atlantic hydrographic section at 26.5°N. Full depth CTDs, with water sampling, LADCP, underway, ship's navigation, ADCP and echo sounding will be obtained. It is expected that similar measurements will be obtained at each mooring site during deployment and servicing periods. Final

recovery, in the context of RAPID, may be in 2008 and it is expected that another trans-Atlantic hydrographic survey will be carried out, subject to SOC core funding.

The Bill Johns collaboration will provide regular hydrographic and velocity profiles and underway data along the Florida Strait cable 4 times per year. These data will augment the cable measurements, which extend back to 1980. The PI wishes to obtain all of these data, in particular the cable data in both raw and calibrated form. It is expected that these data will also be included in the RAPID dataset, but the processing and quality control will already have been carried out and documented by the originators.

Two of the instruments, McLane Moored Profilers (MMPs), will have satellite telemetry, providing real-time data. Development is planned of a 'live' web site of data coming from the moorings, in the style of the TOGA/TAU array. Other data that will be used include wind stress and altimetry, which are available within SOC. Opportunistic hydrographic data, such as the Miami XBT section to Gibraltar, will ultimately be available through BODC, as other RAPID projects also require them. Although modelling work will be carried out to assess the effectiveness of the monitoring array, it is not expected that any output will result that will be used by other projects or included in the final product.

**RDC tasks**

**Ship data**

UK cruises

- Acquire processed CTD, LADCP, underway, SADCP and water bottle analyses data.
- Load bottle data to database.
- Transfer all series to BODC format.
- Visually screen the data for spikes of non-oceanographic origin and apply flags.
- Load to database.
- Create detailed documentation about the instruments, processing, calibrations, quality control procedures and data quality.
- Advertise availability of dataset.

**Moored instrument time series**

- Acquire processed, calibrated 60 current meters, 232 CTD, 76 BPR, 8 ADCP, 8 HOMER, 12 MMP series over span of the project (15 CM, 58 CTD, 2 HOMER, 19 BPR, 2 ADCP, 3 MMP per year).
- Transfer series to BODC format.
- Visually screen the data for spikes of non-oceanographic origin and apply flags.
- Prepare detailed documentation about the sites, moorings, instruments, processing, calibrations, quality control procedures and data quality.
- NODB banking and advertise availability of dataset.

**Real time data**

- Develop operational system to acquire telemetered data and display on web.

**Third Party data**

- Altimetry along array lines will be required. Archive and serve in originator's format.
- Acquire Florida Straits cable data (daily Sverdrup values) from 1980s to present.
  - Transfer to BODC format.
  - Visually screen the data for spikes of non-oceanographic origin and apply flags.
  - Create detailed documentation about the instruments, processing, calibrations, quality control procedures and data quality.
  - NODB banking.
- Bank Johns Western Boundary Array (WBA) servicing cruises:
  - Acquire processed, calibrated CTD, LADCP, underway, SADCP and water bottle analyses data.
  - Load bottle data to database.
  - Transfer all series to BODC format.
  - Visually screen the data for spikes of non-oceanographic origin and apply flags.
  - Load to database.
  - Create detailed documentation about the instruments, processing, calibrations, quality control procedures and data quality.

- Acquire other US data: Bill Johns cruises along 70°W-78°W (Abaco, semi-annually), quarterly Florida Straits cruises, 20 total (CTD/LADCP/SADCP), 52 ADCP transects per year of Florida Strait, Miami-Gibraltar XBT sections.
    - Acquire processed XBT, CTD, LADCP, underway, SADCP, water bottle samples for all cruises.
    - Acquire VOS quarterly Miami-Gibraltar XBT sections ~200 per section = 800 per year.
    - Archive these data in originators format and prepare documentation.

**Major assumptions**

- US proposed field activities take place as scheduled in kick-off meeting presentation.
- US data for archiving are from 2003 up to 2008, when the array is recovered.
- The WBA 1st deployment and recovery cruises will be on UK ships, so there are 2 US WBA servicing cruises, which will be fully banked, and 6 interim Abaco cruises which will be archived.
- All US data will be supplied in a calibrated, documented, quality controlled format 1 year after collection.

**NB:**
**1)** The 6 months BODC staff time to set up protocols and methods to process data from the MOC mooring array in 2005/06 is included as an option and contributes to the maximum total in the costing spreadsheet.

**2)** Staff effort is required beyond formal end of project due to array retrieval and cruises in early 2008. Data processing and calibrations work will have to be carried out before the data can be banked, hence the RDC expect that the data will be submitted sometime in the latter part of 2008.

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC task |
|---|---|
| 2003/04 | Realtime website devt. |
| 2004/05 | Realtime website devt. |
| | Bank UK deployment cruise |
| | Bank UK trans-Atlantic cruise |
| | Archive US data (2003):<br>5 sections of 200 XBT<br>1 Abaco cruise (mid-2003)<br>52 Florida Current ADCP transects<br>4 Florida Strait cruises (2003) |
| | Bank historical Cable data (1980-2003) |
| 2005/06 | Bank UK servicing cruise |
| | Bank UK time series |
| | Archive US data (2004):<br>4 sections of 200 XBT<br>1 Abaco cruise (mid-2004)<br>52 Florida Current ADCP transects<br>4 Florida Strait cruises (2004) |
| | Bank 1 US WBA servicing cruise (Jan 2005) |
| | Bank US Cable data 1 yr (2004) |
| | BODC assist for 6 months with data processing |
| 2006/07 | Bank UK servicing cruise |
| | Bank UK time series |
| | Archive US data (2005):<br>4 sections of 200 XBT<br>2 Abaco cruises (2005/6)<br>52 Florida Current ADCP transects<br>4 Florida Strait cruises (2005) |
| | Bank US Cable data 1 yr (2005) |
| 2007/08 | Bank UK servicing cruise |
| | Bank UK time series |

| | |
|---|---|
| | Archive US data (2006):<br>4 sections of 200 XBT<br>2 Abaco cruises (2007)<br>52 Florida Current ADCP transects<br>4 Florida Strait cruises (2006) |
| | Bank Cable data 1 yr (2006) |
| | Bank 1 US WBA servicing cruise (mid-2006) |
| 2008/09 | Bank UK trans-Atlantic cruise |
| | Bank UK servicing cruise |
| | Bank UK time series |
| | Archive US data (2007):<br>4 sections of 200 XBT<br>52 Florida Current ADCP transects<br>4 Florida Strait cruises (2007) |
| | Bank US Cable data 1 yr (2007) |

## NER/T/S/2002/00482 – Cromwell

**Project data summary**
This is a 1 year feasibility study that will not require data centre effort. If the study is extended data management resources will be sought from the 2nd AO data management budget.

**RDC tasks**
None in duration of 1st AO data management allocation.

## NER/T/S/2002/00484 – Andy Watson

**Project data summary**
This is a pilot study from which no useable data are anticipated by the PI. If the pilot is successful then data management resources will be sought from the 2nd AO data management budget.

**RDC tasks**
None in duration of 1st AO data management allocation.

## NER/T/S/2002/485 – Chris Hughes

**Project data summary**
Time series will be obtained from two mooring arrays deployed in July 2004-2008 at Grand Banks and Halifax, and from instruments incorporated into an NSF-funded mooring array deployed by John Toole in April 2004. The Hughes arrays will be serviced every 2 years, but no plans have been made to perform calibration casts and this needs to be addressed if the data are to be of good quality. The Toole array will be serviced annually.

Ancillary data required by the project includes time series and cruises returned from other locations on the Toole line, and altimeter data that tracks along the 3 lines.

**RDC tasks**
**Ship data**
At present no CTD, water bottle, underway, SADCP sampling is planned. However, BODC has raised the issue with the Science Co-ordinator that the moored instruments (HOMER, SeaCats, MMP) will need some form of calibration on deployment and recovery. It is therefore expected that there will be CTD calibration casts carried out on the UK cruises, but that no LADCP, SADCP or underway measurements will be made.

- Acquire unprocessed CTD data and salinity bottle results.
- Load salinity bottle data to database.
- Transfer all series to BODC format.
- Visually screen the data for spikes of non-oceanographic origin and apply flags.
- Load to database.
- Determine calibrations from water bottle samples and apply.
- Create detailed documentation about the instruments, processing, calibrations, quality control procedures and data quality.
- Advertise availability of dataset.

**Moored instrument time series**

- Acquire fully processed and calibrated BPR time series with documentation detailing all the collection, processing and calibration steps, and unprocessed HOMER, MMP and SEACAT series.
    - Grand Banks and Halifax sections with a total of 24 BPR/IES and 24 HOMER (12 HOMER and 12 BPR in turnaround 2006, and 12 each on recovery in 2008).
    - Hughes instruments on Hughes/Toole array deployed 2004, recovered 2008, yearly turnaround, returning 4 HOMER, 24 BPR/IES, 12 MMP, 84 SEACAT CTD series in total (1 HOMER, 6 BPR/IES, 3 MMP, 21 SeaCat CTD each year).
- Processed data from Toole's instruments on Hughes/Toole array, estimated to be 11 VACM, 1 ADCP, 3 SeaCat CTD or 3 MMP, 2 CM, 1 ADCP.
- Non-profile time series (e.g. SEACAT) will need calibrations determined and applied before transfer.
- Transfer all the series into BODC format.
- Visually screen the data for spikes of non-oceanographic origin and apply flags.
- For unprocessed profile series, determine calibrations and apply.
- Prepare detailed documentation about the site, mooring, instruments, processing, calibrations, quality control procedures and data quality.
- NODB banking and advertise availability of dataset.

**Third Party data**

- Acquire altimeter data that tracks along Grand Banks, Halifax and WHOI sections (2004-2008).
    - Archive as accession and prepare metadata descriptor.
- Acquire US joint array deployment/servicing/recovery cruises as processed, calibrated CTD, LADCP, underway, SADCP and water bottle analyses data.
    - Load bottle data to database.
    - Transfer all series to BODC format.
    - Visually screen the data for spikes of non-oceanographic origin and apply flags.
    - Load to database.
    - Create detailed documentation about the instruments, processing, calibrations, quality control procedures and data quality.
- Acquire remaining John Toole mooring and hydrographic data.
    - 4 cruises with CTD, LADCP, underway, SADCP, water bottles.
    - Additional time series on the WHOI line belonging to Toole: 4 deployments (2004-2008) of MMP
    - Archive data in originator's format and prepare documentation.

**Major assumptions**

- UK cruises have CTD calibration casts but no SADCP, LADCP or underway measurements.
- UK arrays will be recovered in July 2008.
- US proposed field activities take place as scheduled in kick-off meeting presentation.
- All US data will be supplied in a calibrated, documented, quality controlled format 1 year after collection.
- US are processing data from Hughes instruments on joint array and will submit to RDC approximately 1 year after data download.

**NB:**

**1)** If UK arrays are recovered in July 2008, it is expected that processed data will not be ready for banking until late in FY 2008/09, and **some banking effort may occur in FY 2009/10**.

**2)** Similarly, if US data are not available until 1 year post acquisition, then **banking effort may be taking place during FY 2009/10**

**RDC milestones (approximate) based on the points listed above:**

| Year | RDC task |
|---|---|
| 2004/05 | Archive 1 year altimeter data |
| | Bank UK deployment cruise |
| 2005/06 | Archive 1 year altimeter data |
| | Bank US deployment cruise (04/2004) |
| | Archive US hydrographic cruise (10/2004) |
| 2006/07 | Archive 1 year altimeter data |
| | Bank UK servicing cruise |
| | Bank UK time series |
| | Bank US servicing cruise (04/2005) |
| | Bank US/UK joint array time series: Hughes (2004) |
| | Bank US/UK joint array time series: Toole (2004) |
| | Archive US hydrographic cruise (10/2005) |
| 2007/08 | Archive 1 year altimeter data |
| | Bank US servicing cruise (04/2006) |
| | Bank US/UK joint array time series: Hughes (2005) |
| | Bank US/UK joint array time series: Toole (2005) |
| | Archive US hydrographic cruise (10/2006) |
| 2008/09 | Bank UK recovery cruise |
| | Bank UK time series |
| | Bank US servicing cruise (04/2007) |
| | Archive US hydrographic cruise (10/2007) |
| | Bank US/UK joint array time series: Hughes (2006) |
| | Bank US/UK joint array time series: Toole (2006) |
| 2009/10 | Bank US recovery cruise (04/2008) |
| | Bank US/UK joint array time series: Toole (2007) |
| | Bank US/UK joint array time series: Hughes (2007) |