RAPID-WATCH Programme

Data Management Plan

1. Introduction

NERC requires all Directed 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-WATCH 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. Programme budgets include the funds necessary for within project data management for the life of the project. An integral part of the Data Management Plan is an obligation upon RAPID-WATCH 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.

The data policy as defined by the RAPID-WATCH Programme Advisory Group (PAG) is outlined in Annex I.

This plan has been formulated following a review of the specified resource requirements and outputs set out in the RAPID-WATCH Work Plan (online at http://www.noc.soton.ac.uk/RAPID-WATCH/rw/) and a series of discussions between BODC/BADC and the project PIs in order to assess the scale of data production. The details of data involved and timetables for archive of data from all the projects at the data centres are given in Annex III for BODC and Annex IV for BADC.

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

As was the case in RAPID, submission of and access to data will continue to be through a common 'portal' and for the purposes of RAPID-WATCH the term RAPID-WATCH Data Centre (RDC) will refer to BADC and BODC. Data management costs have been allocated in the RAPID-WATCH budget for RDC services.

The basis of data management for the individual projects is outlined in Annex III for data resulting from fieldwork and Annex IV for model output.

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-WATCH participants can easily request available data. The RDC will service data requests by RAPID-WATCH participants and it is expected that automatic data download for observations series will be available from autumn 2009.

Following the completion of RAPID-WATCH 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. NERC's policy towards exploiting and making data available to third parties must be agreed at the outset.
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-WATCH data sets managed by BODC and BADC. Nevertheless, PIs need to be aware of this framework.

4. Data Acquisition

RAPID-WATCH data cover oceanographic 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.

A well structured and user-friendly identification system is essential for cruise-based data collection and sample labelling. Such arrangements are the responsibility of the cruise Principal Scientist. Station identifiers, navigational information and "basic" oceanographic data must be provided to the RDC by the Principal Scientist as soon as possible after a cruise. 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 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.

5. 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-WATCH 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.

6. 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 offered to the RDC in the recommended CF compliant NetCDF format, although there may be exceptions (particularly PP and HDF will also be accepted in extreme situations). Documentation on formats and conventions is available from the RDC (http://www.badc.rl.ac.uk/formats/), which also provides links to numerous downloadable free software packages to support NetCDF file handling.

Submission of field data continue as in RAPID, mainly accessed via the computer network. FTP is recommended for the submission of large amounts of model output. At an early stage Investigators should discuss the options for model data submission with the RDC.

7. 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.

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

Annex I: RAPID-WATCH data policy

An important aim of RAPID-WATCH is to ensure that the data from the RAPID observing system is made available to the wider climate change science community as soon as possible after collection. To facilitate this, the following data policy has been recommended by the Programme Advisory Group (PAG) in discussion with PIs of the RAPID-WATCH observing system and PIs of projects funded under the data exploitation AO, and agreed by the Project Executive Board (PEB). It will apply to data[1] from all projects funded through RAPID-WATCH:

- Data from the RAPID-WATCH observing system should be lodged with the RAPID-WATCH Data Centre (RDC) as soon as possible and in general no later than 6 months after acquisition[2], together with such metadata as are defined under the RAPID-WATCH Data Management Plan (DMP).
- Model output and data-model syntheses deemed to be of wider interest should be lodged with the Data Centre according to agreed schedules[3], and no later than the end of the project,
- Free and open access[4] to all RAPID-WATCH data will be available to anyone accepting the terms and conditions for data use via the Data Centre web portal.
- Users of RAPID-WATCH data are required to acknowledge RAPID-WATCH in any published work making use of the data. Within 3 years of the data being collected users are also expected to acknowledge the PI and/or co-workers (as appropriate) in any resulting papers.
- The RAPID-WATCH data policy will apply to all projects funded by NERC[5] under RAPIDWATCH (2008-2014), and to all data from the observing system at 26 North and the WAVE arrays in the Deep Western Boundary Current (DWBC), including data from the period funded by RAPID (2001-2007). Data from other projects funded under RAPID will continue to be subject to the RAPID data policy.
- PIs and/or co-workers failing to comply with the RAPID-WATCH data policy would be subject to appropriate sanctions.

NOTES

1. RAPID-WATCH data includes all observations from the RAPID-WATCH observing system at 26 North and in the DWBC (including data acquired during RAPID 2001-2007); model output from projects funded under the RAPID-WATCH Exploitation AO and deemed to be of wider interest; data syntheses and data-model syntheses carried out as part of projects funded by RAPID-WATCH.

2. As soon as possible after acquisition: the date of acquisition is the date on which the data was downloaded from an instrument; in the case of instruments recovered during a cruise, the acquisition date will be the end-date of the cruise. The time-scale for lodging data with the RDC may vary between data types; some data (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.

All data from measurements are to be calibrated and banked within 6 months, but in exceptional circumstances (agreed in advance with the PEB) data may be submitted later than this. This is to ensure that data acquired are available to the RAPID-WATCH and wider climate change communities on a time scale that allows data use to be considered as part of the observing system review in 2011.

3. Schedules for delivery of model output agreed between the Data Centre, the PIs and the PEB will be linked to project milestones to ensure that where possible model output and model-data syntheses may be taken into consideration as part of the observing system review in 2011.

4. Free and open access: unrestricted access for any use, including academic, non-profit and commercial, free of any charge except the cost of data provision.

5. The RAPID-WATCH data policy will be applied to projects funded by NERC: PIs of collaborating projects in the USA and Canada will not be expected to comply with the terms of this data policy, but will be encouraged to do so, and the RDC will provide the necessary support to make this possible.

Annex II: RAPID-WATCH 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-WATCH 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 be as comprehensive as possible, and include information such as 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. Where non-self describing files are submitted, metadata may also include information on the format in which the data are stored, and the order of the variables, to allow potential users to access them. Metadata pertaining to software models should also 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).

2. Metadata for tables of numbers (observations or model output)

Ideally, each data file should include a header containing the metadata. If the file is not self describing and 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.

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

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.

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

- 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.

3. Metadata for software

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.

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.

4. 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-WATCH archive documentation directory. These documents may for example include technique description, possible use of the data, study conclusions, etc.

Annex III: RAPID-WATCH Observations data banking

1. Introduction

RAPID-WATCH observations will be collected from 2008 to 2014. The Hughes and Cunningham MOC projects will comprise the observational component of RAPID-WATCH and all their data will be banked at BODC. Modelling projects will be catered for at the BADC, as per RAPID.

As the MOC and WAVE arrays are in position until mid-2008 under RAPID, it is expected that data collection under RAPID-WATCH will start at the former planned recovery time, ie on turnaround in 2008.

The plan for management of data from the MOC and WAVE projects was produced from information on fieldwork and deployments listed in the RAPID-WATCH work plan. The plan encompasses all calibrated data arising from fieldwork. The profiling of data management staff effort assumes that data submission will occur no later than 1 year after collection (i.e. downloading from the instrument logger). Thus in the case of mooring arrays the data would be expected to be submitted to the data centre no later than one year after turnaround of the instruments on the array. Thus data banking for RAPID-WATCH is expected to continue into 2015. However, if data submission is delayed then banking effort will be likely to extend into the 2016/17 financial year. Irrespective of when data are submitted, the data centre priority will be to get those data banked and available as soon as possible.

Ultimately, the success of RAPID-WATCH is dependent on making data available for use in as short a time as possible, ideally within a few months of instrument turnaround. However, this is heavily dependent on the scientists and collaborators involved.

Both projects have clearly defined US collaborators, and it is expected that these collaborations will continue as they are throughout the RAPID-WATCH period. A detailed breakdown for the Hughes/Toole and Cunningham/Johns collaborations of expected fieldwork and BODC effort over the RAPID-WATCH period is given in the following sections.

2. Continuation of NER/T/S/2002/485 RAPID WAVE

Under RAPID, the aim was to obtain time series from three mooring arrays: two RAPID lines deployed in July 2004-2008 at Grand Banks and Halifax; and one NSF-funded mooring array - Line W (incorporating UK instruments) deployed by John Toole in April 2004. However, the RAPID lines experienced heavy instrument losses and were consolidated to a single line.

Under RAPID-WATCH, the consolidation will remain as a single line of moorings off Halifax and the work will be carried out in partnership with the Bedford Institute of Oceanography (BIO). Line W will include POL BPRs in the line moorings and this Toole array will be serviced annually. An outline of cruises for RAPID WAVE is given in Table 1. All data from the arrays and servicing cruises will be fully banked and an estimated timescale for banking tasks is shown in Table 2.

All Line W and Halifax BPR data will be processed at POL. BIO will carry out the processing and calibration of all Halifax line moorings and CTDs. John Toole's group will continue to process and calibrate their line mooring and ship data. It is assumed that they will continue to make this available via their public access website. Note that the timescale tends to be variable.

Ancillary data required by the project under RAPID included time series and cruises returned from other locations on the Toole line. It is assumed that these are still required. These data will not undergo full banking; instead they may be archived in their existing format and documentation describing them will be produced and stored alongside.

RDC tasks

Ship data (NB: these are third party data)

10 array servicing cruises (BIO & WHOI)

- 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
(numbers of instruments from RAPIDWATCH work plan and LineW website)

- Acquire fully processed and calibrated time series
  - Halifax line 6 BPRs, yearly turnaround (POL processed)
  - Halifax line 6 CTD moorings, yearly turnaround (BIO processed)
  - Halifax line 5 ADCP moorings, yearly turnaround (BIO processed)
  - Line W 6 BPRs, yearly turnaround (POL processed)
  - Line W MMP moorings (3 MMP, 6 CTD, 6 VACM), yearly turnaround (WHOI processed)
  - Line W VACM moorings (21 CTD, 10 VACM), 2 yearly turnaround (WHOI processed)
- 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 of dataset.

Third Party data

- Acquire WHOI hydrographic survey cruises as processed, calibrated CTD, LADCP, underway, SADCP and water bottle analyses data.
  - Archive data in originator's format and prepare documentation.

Major assumptions

- BIO cruises have CTD calibration casts but no SADCP, LADCP or underway measurements.
- Arrays will be recovered in mid 2013.
- WHOI/BIO proposed field activities take place as scheduled in RAPID-WATCH work plan and Line W website.
- All WHOI/BIO 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.

Table 1: Outline of WAVE cruises from RAPID-WATCH work plan

Year	Line	Ship	BRP moorings	BRP/CTD moorings	ADCP moorings	Mooring operations	Sampling stations	Data expected
2009	Halifax	Hudson AZMP-BIO	0	6	5	Deployment	6	
2009	W	Oceanus W-WHOI	6	0	0	Recovery/redeployment	0	2010
2010	Halifax	Hudson AZMP-BIO	0	6	5	Recovery/redeployment	6	2011
2010	W	Oceanus W-WHOI	6	0	0	Recovery/redeployment	0	2011
2011	Halifax	Hudson AZMP-BIO	0	6	5	Recovery/redeployment	6	2012
2011	W	Oceanus W-WHOI	6	0	0	Recovery/redeployment	0	2012
2012	Halifax	Hudson AZMP-BIO	0	6	5	Recovery/redeployment	6	2013
2012	W	Oceanus	6	0	0	Recovery/redeployment	0	2013

		W-WHOI						
2013	Halifax	Hudson AZMP-BIO	0	6	5	Recovery	6	2014
2013	W	Oceanus W-WHOI	6	0	0	Recovery	0	2014

Table 2: Timetable of WAVE data banking tasks at BODC

Financial year	Fieldwork	Banking task	Comments
2007/08	BIO servicing cruise[1]		RAPID-funded[1]
2008/09	BIO servicing cruise** WHOI servicing cruise[1] WHOI hydro. cruise	Bank 2007 BIO Hudson servicing cruise[1] Bank UK time series from 2007 recovery[1]	RAPID-funded[1] ** Included in RAPIDWATCH DM costs as 2007 cruise not originally accounted for in RAPID.
2009/10	BIO servicing cruise WHOI servicing cruise WHOI hydro. cruise	Bank 2008 BIO Hudson servicing cruise** Bank UK+BIO time series from 2008 recovery[1] Bank 2008 WHOI servicing cruise[1] Bank WHOI time series from 2008 recovery[1] Archive 2008 WHOI hydro. cruise	
2010/11	BIO servicing cruise WHOI servicing cruise WHOI hydro. cruise	Bank 2009 BIO Hudson servicing cruise Bank UK+BIO time series from 2009 recovery Bank 2009 WHOI servicing cruise Bank WHOI time series from 2009 recovery Archive 2009 WHOI hydro. cruise	
2011/12	BIO servicing cruise WHOI servicing cruise WHOI hydro. cruise	Bank 2010 BIO Hudson servicing cruise Bank UK+BIO time series from 2010 recovery Bank 2010 WHOI servicing cruise Bank WHOI time series from 2010 recovery Archive 2010 WHOI hydro. cruise	
2012/13	BIO servicing cruise WHOI servicing cruise WHOI hydro. cruise	Bank 2011 BIO Hudson servicing cruise Bank UK+BIO time series from 2011 recovery Bank 2011 WHO servicing cruise Bank WHOI time series from 2011 recovery Archive 2011 WHOI hydro. cruise	
2013/14	BIO recovery cruise WHOI recovery cruise WHOI hydro. cruise	Bank 2012 BIO Hudson servicing cruise Bank UK+BIO time series from 2012 recovery Bank 2012 WHOI servicing cruise Bank WHOI time series from 2012 recovery Archive 2012 WHOI hydro. cruise	
2014/15		Bank 2013 BIO Hudson recovery cruise Bank UK+BIO time series from 2013 recovery Bank 2013 WHOI recovery cruise Bank WHOI time series from 2013 recovery Archive 2013 WHOI hydro. cruise	

3. Continuation of NER/T/S/2002/00481 RAPIDMOC

Under RAPID, mooring arrays, covering western, eastern and mid Atlantic sections, were deployed over a 4 year period, 2004-2008. The western array has been serviced annually by US ships, in collaboration with NSF-funded Bill Johns. UK cruises have been servicing the mid-Atlantic and eastern arrays 6 monthly to annually. This will continue under RAPID-WATCH from 2008-2014. All data from the arrays and servicing cruises will be fully banked.

Also, there will be a full trans-Atlantic hydrographic section in 2014. This will be NOC core-funded, but the data will form part of the RAPID final data set. (The NOC core-funded 2004 and 2008 trans-Atlantic sections were included in the RAPID data set under RAPID. The 2008 section is now taking place in 2009.)

Ancillary data will be archived as per RAPID, including regular hydrographic and velocity profiles and underway data along the Florida Strait cable 4 times per year and data from the Abaco cruises.

An outline of the main banking tasks is given in Table 3. Note: There have been developments in using gliders to form part of the RAPID-WATCH data set. The glider data have not been included in Table 3 as it is unknown how much glider data will result and when it might be submitted for banking. However, BODC will include it as part of the banking process as and when it arrives.

RDC tasks

Ship data

Array servicing 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

- Process and calibrate moored CTD data from entire array as per RAPID.
- Accession processed, calibrated current meter, BPR and moored CTD data.
- 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.

Glider data

- Acquire processed, calibrated glider data.
- 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

- Continue to manage and support the real time website.

Third Party data

- Altimetry along array lines will be required. Archive and serve in originator's format.
- Acquire Florida Straits cable data yearly.
  - 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 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: Molly Baringer's 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 data for archiving are from 2003 up to 2014, when the array is recovered.

- US WB moorings are serviced every 18 months. UK moorings are serviced annually. Assume that UK will always have two cruises per year. US moorings will sometimes coincide with UK cruise, other times it will be separate cruise.
- All US data will be supplied in a calibrated, documented, quality controlled format 1 year after collection.
- Glider data is lodged at BODC in a fully processed/calibrated format.

NB:
1) BODC staff effort is required beyond formal end of project due to array retrieval and cruises in early 2014. 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 2014, which will involve banking effort into 2015.

Table 3: Timetable of RAPIDMOC data banking tasks at BODC, generated using cruises listed in RAPID-WATCH work plan.

Financial year	Fieldwork	RDC task	Comments
2008/9	Spring US/UK WB servicing cruise (A) Autumn US hydro cruise (B) Autumn UK EB/MAR servicing cruise (C)	Bank US cable data (2008)[1] Bank (A) 2008 Spring WBcruise[1]	RAPID-funded[1]
2009/10	Spring UK WB servicing cruise (D) Autumn UK EB/MAR servicing cruise (E) Autumn US WB servicing cruise (F)	Archive (B) 2008 US hydro cruise Bank (C) 2008 Autumn EB cruise Archive 2008 US third party data Bank US cable data (2009) Bank (D) 2009 UK spring cruise	RAPID-funded[1]
2010/11	Spring US WB servicing cruise (G) Autumn US hydro cruise (H) Autumn UK EB/MAR servicing cruise (I)	Bank (E) 2009 Autumn UK EB cruise Archive (F) 2009 Autumn US hydro Bank (G) 2010 Spring servicing cruise Bank 2008-2009 array data Bank US cable data (2010) Archive 2009 US third party data	
2011/12	Spring US WB servicing cruise (J) Autumn US hydro cruise (K) Autumn UK EB/MAR servicing cruise (L)	Bank (H) 2010 Autumn US WB cruise Bank (I) 2010 UK Autumn EB cruise Bank (J) 2011 US Spring WB cruise Bank 2009-2010 array data Bank US cable data (2011) Archive 2010 US third party data	
2012/13	Spring US WB servicing cruise (M) Autumn US WB servicing cruise (N) Autumn UK EB/MAR servicing cruise (O)	Archive (K) Autumn US hydro cruise Bank (L) 2011 Autumn UK EB cruise Bank (M) 2012 Spring US WB cruise Bank 2010-2011 UK array data Bank US cable data (2012) Archive 2011 US third party data	
2013/14	Spring US WB servicing cruise (P) Autumn US hydro cruise (Q) Autumn UK EB/MAR servicing cruise (R)	Bank (N) 2012 Autumn US WB cruise Bank (O) 2012 Autumn UK EB cruise Bank (P) 2013 Spring US WB cruise Bank 2011-2012 array data Bank US cable data (2013) Archive 2012 US third party data	
2014/15	Spring UK WB/MAR/EB recovery cruise[2] (S)	Archive (Q) 2013 US hydro cruise Bank (R) 2013 Autumn UK EB cruise Bank (S) 2014 Spring servicing cruise Bank 2012-2013 array data Bank US cable data (2014) Archive 2013 US third party data	Trans-Atlantic section[2]
2015/16		Bank 2013-2014 array data	

Annex IV: RAPID-WATCH model output archive

1. Introduction

This Annex describes the model data archival support for RAPID-WATCH requested by the projects involved. The original provisional costing the RDC provided to NERC some time ago for RAPID-WATCH was based on the BADC providing support for 4 modelling projects, each requiring the storage of 2 TB each (8TB in total). In the end, 5 modelling projects have been funded, and these are detailed below. Abstract sections below are taken directly from NERC's Grants on the Web (GotW).

The smaller number of projects for RAPID-WATCH (compared with RAPID) should mean that it is easier to build good relationships with the data providers and ensure that the RDC is "embedded" into the projects from an early stage.

A scoping study was carried out by BADC to gather information about the data that would be produced by each project. For some of the projects, this process was limited as they had only recently commenced, and some of the implementation details of the projects (such as finalised project management plans and recruitment) had yet to be resolved.

2. Model data archive requirements

A summary of the data storage requirements learned from the scoping study is shown in Table 4.

Table 4: Summary of the data storage requirements for RAPID-WATCH model projects

Project	**PI**	**Storage requested (Tb)**
Value Of the Rapid array for climate predictions (VALOR)	Rowan Sutton	20
Change in the Atlantic Atmosphere Ocean System: ChAAOS	Myles Allen	24
RAPID-RAPIT	Peter Challenor	50
What are the roles of natural and human drivers in historical changes in the Atlantic Meridional Circulation?	Simon Tett	1
Meridional Overturning circulation at 26N and the North Atlantic heat COtent (MONACO)	Neil Wells	0.01
	Total	95

A total of approximately 95 Tb is estimated to be required to support RAPID-WATCH modelling activities. These data will be curated at the BADC beyond the end of the programme, and continue to be available under the terms of the RAPID-WATCH data policy. Note also that for Peter Challenor's project, the availability of this storage is essential for the project to proceed as planned. The timeline for archive of model output is shown in Table 5.

Table 5: Timeline for archive of data from RAPID-WATCH model projects

	Q1	Q2	Q3	Q4
2009-10	RAPID WATCH Scoping study.	.	Complete RAPID-WATCH DMP. Set up RAPID WATCH dataset at BADC.	Archive initial data from Challenor project.
2010-11	Archive data from Challenor project.	Archive data from Challenor project.	Archive initial data from Sutton Project.	Archive data from Challenor project.
2011-12	Archive initial data from Wells project.	Archive data from Challenor project.	Archive initial data from Tett project.	Archive data from Challenor project.
2012-13	Archive final data from Sutton project.	Archive final data from Tett project.	Archive final data from Wells project.	Archive final data from Allen project.
2013-14	Archive data from Challenor project.	Archive data from Challenor project.	Archive final data from Challenor project.	
2014-15	Complete collection of RAPID WATCH data.		All RAPID WATCH model data available behind on-line interfaces.	

An overview of each of the RAPID-WATCH modelling projects follows in the next sections.

3. VALue Of the Rapid array for climate predictions (VALOR)

NERC Reference: NE/G007845/1

Principal Investigator: Rowan Sutton

Abstract: The Atlantic Meridional Overturning Circulation (AMOC) involves a northward movement of warm surface waters balanced by a southward movement of cold deep waters. The net effect is to transport ~1PW of heat northwards. This heat is released to the atmosphere in mid-high latitudes, where it acts to warm the climate, notably in northern Europe. The future behaviour of the AMOC is an issue of major importance in climate prediction. Forecasts presented by the Intergovernmental Panel on Climate Change (IPCC) suggest that, in response to greenhouse gas forcing, the AMOC may slow down, reducing the northward transport of heat by the Atlantic Ocean, and leading to a cooling of northern Europe that could offset anthropogenic

warming. Moreover, there is evidence from palaeoclimate records that the AMOC can undergo very rapid transitions such as a total shutdown within little more than a decade. It is possible that increasing levels of greenhouse gases could trigger such a rapid change with potentially serious consequences for societies in Europe and other regions surrounding the Atlantic basin. In the face of such risks, there is an obvious need for better, more quantitative, forecasts of the future behaviour of the AMOC. Such forecasts could provide early warning of possible rapid changes in the AMOC in future The RAPID array is a measurement system for observing the current state of the AMOC. Established in 2004, the array is providing direct and continuous measurements of the AMOC for the first time, and under the RAPID-WATCH programme these measurements will be maintained for a decade. The overarching goal of the VALOR project is to assess the value of these observations for predicting the future behaviour of the AMOC, and its impacts on climate. The project will explore a range of issues concerning the design of a potential AMOC prediction system. To achieve its goals VALOR will exploit the RAPID observations in a variety of ways. First the observations will be used as independent data to assess the quality of current ocean "analyses" (An "analysis" provides a quantitative description of the state of the ocean at a given time.). Next, they will be used to improve the analyses. Finally, they will be used to provide the starting conditions for a large set of "hindcasts". Hindcasts are predictions made from a date in the past, which only make use of information that would have been available at that date. These predictions can then be compared to what actually happened to assess prediction skill. VALOR will carry out a suite of hindcast experiments to quantify the extent to which the RAPID observations can improve the skill of predictions of the AMOC and its impacts on climate. An important dimension of the project is that it involves agencies who are directly involved in operational climate forecasting: the Met Office Hadley Centre (MOHC) and the European Centre for Medium Range Weather Forecasting. By involving these partners from the start the project will benefit from their experience and expertise, and the scientific advances achieved through the research will feed directly into better climate predictions.

Project data summary
The project started in May 2009 and will run for 3 years.

Data will be delivered in 2x 10Tb parts. In the first stage, ocean synthesis data will be produced. The second phase will see the production of model hindcasts, using NEMO01, NEMO025 and the UM. The reanalysis data will produce monthly and 5 day average fields (4Tb and 20 Tb respectively). Data will be converted to NetCDF, and the first data should be ready by summer 2010.

Some of the synthesis data will use ECMWF and MOHC data, so there may be licensing issues which need to be investigated. 3rd party data: Bryden section data (from the BODC) may be of use.

Data archival storage requested: 20Tb

4. Change in the Atlantic Atmosphere Ocean System: ChAAOS

NERC Reference: NE/G007799/1

Principal Investigator: Myles Allen

Abstract: This project uses two novel inputs to address the fundamental problem of understanding observed variability and change in the Atlantic Ocean in the context of the global coupled atmosphere-ocean system. Using a combination of large-ensemble perturbed-physics experiments made possible with distributed computing and new adjoint-based estimates of the recent ocean state together with uncertainty therein, we aim to identify free-running versions of an atmosphere-ocean general circulation model that, for the first time, actually reproduce the full evolution of the large-scale climate system over the past 15 years. The ensemble of such models will provide unique insights into the origins, nature and predictability of recent changes in ocean state, together with a valuable tool for assessing future predictability and the risk of substantial Atlantic Meridional Overturning Circulation changes in the longer term. Coupled models currently used both for decadal prediction and longer-term projections of the response to changing boundary conditions typically rely on the comparison of model anomalies from the model climatology with observed anomalies from some estimate of the "real world" climatology. This is a fundamental problem when either (a) the response to external forcing is uncertain and comparable to any predictability that may arise from the initial state or (b) the system contains significant non-linearities that are likely to impact on any forecast. We will use an entirely novel approach to initialising coupled models directly from a state-of-the-art ocean analysis, using direct perturbation of coupled

model parameters to find model-versions that track the real world over the past 15 years. This very challenging objective is made feasible by the unprecedented computing resources, allowing multi-thousand-member ensembles with a fully coupled atmosphere-ocean general circulation model, provided by distributed computing. Our approach will be to initialize tens to hundreds of thousands of perturbed versions of two AOGCMs from the ECCOc ocean analysis, perturbed to allow for both observational and structural uncertainty, estimated from the discrepancies between ECCOc and other analyses. We will use the statistical techniques of likelihood profiling and importance sampling to identify parameter/analysis combinations that allow models to continue to "shadow" the analysis initially for six months and subsequently, as we home in on promising perturbations, out to the full 15 years. The models used will be HadCM3, which is already set up for distributed computing applications, and a new model based on coupling the HadAM3P model to the MITgcm used in the ECCOc analysis, exploiting information on the parameter-sensitivities of both models that is already available from past ensemble experiments and (in the case of the MITgcm) from the model adjoint. Successful model-versions will then be run free over the full 20th century (for HadCM3) or from 1975 onward (for HadAM3P/MITgcm) to assess the range of AMOC trends they generate in response to total external forcing and anthropogenic forcing only over the past two decades. They will also provide an range of initial conditions for an ensemble prediction experiment to be performed by the VALOR consortium. In addition to its scientific benefits, this project will provide a significant public outreach opportunity, allowing participants to see RAPID data being used directly to address problems of clear and immediate concern.

Project data summary
The project started on 1/6/2009 and will run for 3 years. Hiro Yamazaki is co-ordinating the project at Oxford.

24Tb of model output will be produced for archival at the RDC by this project, with data production taking place throughout the lifetime of the project. These data will be in NetCDF format, though there is some work to do concerning thinking about the best way to store output from CPDN projects to ensure that they are both preservable and accessible. Further details will be required by the RDC

Data archival storage requested: 24Tb

5. RAPID-RAPIT

NERC Reference: NE/G015368/1

Principal Investigator: Peter Challenor

Abstract: North-west Europe has a relatively mild climate in part because of heat pulled north through the Atlantic by the overturning. There is a risk that global warming will cause this circulation to rapidly decrease with consequences involving not only colder winters for Europe but also changes in sea level and precipitation. This project will carry out a risk assessment of rapid changes of the Atlantic overturning. We will use two models of the climate system, HADCM3, the Hadley Centre model used in the IPCC AR4, and CHIME, a global climate model developed at the National Oceanography Centre, Southampton. This has the same atmospheric model as HADCM3 but has a very different structure to the ocean component. Making use of the resources of climateprediction.net we will run a large ensemble of both models to assess the uncertainties in the system. We will then use modern Bayesian statistical techniques to combine model output, data and expert opinion in our risk assessment. An assessment of the utility of the data from the RAPID-WATCH arrays is an important aim of the project.

Project data summary
This project emerged as a result of the RAPID-WATCH ‘Sandpit”, and as such includes a large number of Co-I’s. The project runs for 4 years from 01/05/2009.

Part of the modelling work will be done using Climate prediction.net at Oxford, in collaboration with Myles Allen. The project will produce numerical model output (CF NetCDF) continuously throughout the duration of the project, with delivery commencing at the end of 2009.

It is estimated that between 32 and 64 Tb of storage will be required, and this is essential for the success of the project. To save costs, the BADC archive will provide the main data archive for RAPIT, and as such will be

actively used from the outset. A middle estimate was used for the expected size of the dataset. No 3rd party data are required

Data archival storage requested: 50Tb

6. What are the roles of natural and human drivers in historical changes in the Atlantic Meridional Circulation?

NERC Reference: NE/G007861/1

Principal Investigator: Simon Tett

Project data summary
The project is still at a very early stage and not all of the implementation details have been resolved. The project started in April 2009 and will run for 3 years, although this project has been cut down from the original project proposal and funding has only been secured until 2 years 2 months from project starting.

3rd party data required: Access to both long oceanographic instrumental records and ocean analyses from the result of ocean analysis systems for the North Atlantic region. Most of the data will come from open source databases and through personal contacts.

The project will produce synthesis of several ocean reanalysis datasets and multi-model datasets, and the data will be produced in NetCDF format. The first data are expected to be ready for archive around October 2010. Simon also noted that there may be some issues with licensing of the resultant dataset due to some of the datasets used.

Data archival storage requested: 1Tb

7. Meridional Overturning circulation at 26N and the North Atlantic heat COtent (MONACO)

NERC Reference: NE/G00787X/1

Principal Investigator: Neil Wells

Abstract: RAPID-WATCH is providing a unique continuous observational time series of the Atlantic meridional overturning circulation (MOC) at 26 N for the first time. The crucial question in this proposal is the link between the MOC and meridional heat transport (MHT) from the RAPID-WATCH observing system and the variability of oceanic heat content in the whole of the North Atlantic inferred from Argo floats. We have strong evidence that the ocean heat content variability from seasonal to interannual time scales (months-10 years) is mainly in the upper 2000 m of the ocean, and therefore can be sampled using the Argo floats. Argo is an international experiment to measure the temperature and salinity of the upper 2000 m of the global ocean using over 3000 profiling floats. Understanding the links between the full depth monitoring of the MOC at 26 N and the heat content of the upper 2000m of the North Atlantic is important. For example, a decrease in heat transport at 26 N and a normal heat transport at the latitude of UK, would lead to cooling of the ocean between these two latitudes, which could influence the strength of westerly winds and alter the temperature and rainfall patterns over the UK and Northern Europe.

Project data summary
The project started on 1/6/2009 and has been funded for 12 months, with an expectation of an extension for a further 24 months.

The actual data curation requirements for this project are expected to be very modest. The data will be synthesis maps based on ARGO float data, and will be produced by Feb 2010 at the earliest, with all data ready by May 2012. Data are anticipated to be supplied in ASCII format.

3rd Party data: data from the RAPID WATCH project VALOR will be used.

Data archival storage requested: 0.01 Tb