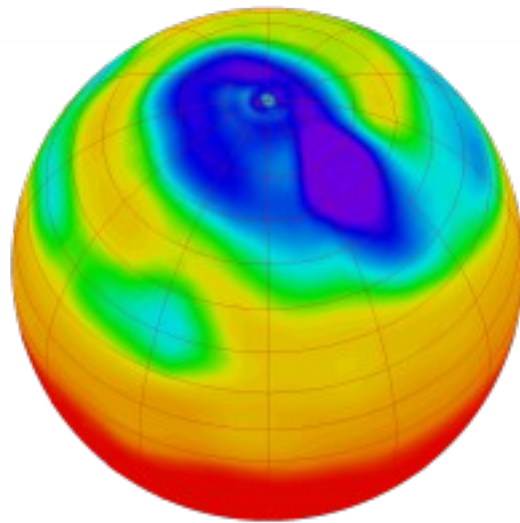




Science & Technology  
Facilities Council



**STFC**

**Centre for Environmental Data Archival**

**Annual Report**

**2009**

**(April 2008-March 2009)**

This short report provides summary information about some of the activities carried out in CEDA during 2008/2009, and introduces plans for 2009/2010. There are three sections to the report, which comprises: (1) an introduction to CEDA itself, the CEDA strategy, and current objectives; (2) a report on 2008/09 itself, which begins with a summary of the goals for 2008/09, has notable events, highlights and “metrics” (including details of funding, dataset usage and publications etc.); and (3) A “forward look” section, outlining the main objectives for 2009/2010.



**British Atmospheric  
Data Centre**

NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE  
NATURAL ENVIRONMENT RESEARCH COUNCIL



**NATURAL  
ENVIRONMENT  
RESEARCH COUNCIL**



**National Centre for  
Earth Observation**

NATURAL ENVIRONMENT RESEARCH COUNCIL



## An Introduction to CEDA

The Centre for Environmental Data Archival is based at the UK Science and Technology Facilities Council's Rutherford Appleton Laboratory (RAL) near Oxford. CEDA staff are members of the RAL Space Science and Technology Department. The director of CEDA is Dr Bryan Lawrence.

*CEDA exists to enable the efficient delivery of environmental science programmes through the provision of effective data and information services.*

In doing so, it delivers the British Atmospheric Data Centre on behalf of the National Centre for Atmospheric Science, and the NERC Earth Observation Data Centre on behalf of the National Centre of Earth Observation, other data support activities (e.g. the Intergovernmental Panel for Climate Change Data Distribution Centre) and carries out appropriate informatics research.

In general this is achieved by:

- Facilitating environmental science by running data and information services.
- Operating efficient data curation services.
- Deploying, and where necessary, developing data service technologies.
- Maintaining close contacts with the research community.
- Maintaining a national capability in data management expertise.
- Contributing to and learning from the international community via knowledge exchange activities and involvement with standards bodies, programme steering boards and external committees.

Our success is measured by:

- The impact of our work on the delivery of environmental science programmes,
- Community recognition of CEDA activities, via public and private acknowledgement, invitations to join external steering boards and committees, the public uptake of CEDA services and the receipt of ongoing funding.
- The publication record of CEDA staff, which includes journal and conference papers for the academic audience, as well as magazine articles and media outreach for the general public.
- Our prime sponsors, the national centres of Earth Observation and Atmospheric Science (NCEO and NCAS respectively) and their prime funder (the Natural Environment Research Council, NERC), are seen as global leaders in data management through CEDA activities, and
- The career development and job satisfaction of CEDA staff.

In the future we hope to further see that those who provide data for management by CEDA receive academic credit for time and energy spent producing quality, well documented data products.

CEDA currently receives major funding from:

1. The Natural Environment Research Council (NERC)
2. The UK Government Department of Food and Rural Affairs (Defra)
3. The European Space Agency
4. The European Commission
5. The UK Joint Information Systems Committee (JISC)

along with a host of smaller contributions.



## Strategic Objectives

### Objectives

- Operate key data centres: the British Atmospheric Data Centre (BADC, <http://badc.nerc.ac.uk>), the NERC Earth Observation Data Centre (NEODC, <http://neodc.nerc.ac.uk>), and the IPCC Data Distribution Centre (<http://ipcc-data.org>).
- Enhance operational data delivery services.
- Support scientific users of atmospheric data and scientists providing data for the archives.
- Provide data management outreach and training. Provide technology and atmospheric science training and outreach.
- Exploit project opportunities to develop potential new services.
- Participate in research programmes in our key areas of expertise, which include atmospheric science, earth observation, geospatial informatics and data management per se,
- Provide general scientific guidance in our areas of expertise.
- Participate in appropriate curation and informatics activities.
- Develop, maintain, and share data management expertise.
- Develop and manage knowledge exchange processes for the benefit of wider communities, both academic and beyond.
- Engage with other data centres nationally and internationally in collaborations to further our objectives.
- Develop people in the CEDA team via a process of continual professional development.

### Methods

#### Science Support – acquiring data and engaging producers and consumers of data

- Build partnership with stakeholder communities and other data archival/curation organisations.
- Exchange knowledge with the wider environmental science community via delivery of appropriate materials, training and workshops.
- Acquire and curate appropriate data from external bodies in accordance with community requirements.
- Support data providers and users via development of data management plans, practical advice, and help desk services, prioritized by requirements of CEDA funders.
- Support NCAS and NCEO by managing data from key facilities (especially FAAM, FGAM and ARSF).
- Support NCAS and NCEO by providing technical and management advice and training for students and existing data activities in their constituent bodies and institutes (especially FAAM and FGAM).
- Provide sufficient technical support to enable users to make appropriate use of our data.

#### General – manage the data archives in accordance with best practice

- Regularly evaluate best practice in the community and how CEDA conforms with it.
- Provide tools and services (e.g. visualisation, subsetting, download) to facilitate the best use of data.
- Document data and services according to best practice, and where necessary provide governance for underlying vocabularies.
- Publicise and promote data products and services via developing and deploying appropriate electronic indexing services and portals.
- Promote fair access to data and information.
- Promote good data management practice by participation in and leadership of, committees, projects, standards bodies and collaborative organisations.
- Ensure that metadata both identifies the “designated community” and is regularly reviewed.
- Meet obligations to make data and information in accordance with relevant legislation (freedom of information, environmental information regulations, INSPIRE<sup>1</sup>).

---

<sup>1</sup>INSPIRE:



- Protect the rights of data providers by formally arranging licences governing the use of data supplied and applying access restrictions where appropriate.
- Work with other data centres, particularly those within NERC, to develop common approaches to data management and service deployment, where such common approaches provide scientific and/or fiscal benefit.
- Establish and improve relationships with key organisations in the private sector, to improve knowledge transfer, and enhance exploitation of private sector data in academia and vice-versa.

#### **Development – design, construct and deploy data handling tools to enhance user experience.**

- Maintain and upgrade the infrastructure and computing systems to support data holdings and data access, and provide backup systems and data as appropriate.
- Evaluate external software products, and where necessary and possible, purchase and/or modify them according to local needs (in doing this, aim to exploit open-source offerings before resorting to paying for software which may have unquantifiable future costs).
- Construct necessary software according to best practice and in such a way that it is most likely to be quickly deployable and maintainable over its planned life cycle.
- Improve interoperability, management and deployability of data services via developing tools which exploit appropriate standards (from de-facto community conventions through de jure standards such as those of the Open Geospatial Consortium and the International Standards Organisation).
- Contribute to the design and evolution and maintenance of relevant standards activities.

#### **Research – in informatics and environmental science as necessary.**

- Carry out active research in both informatics and environmental science as needed to ensure systems and activities can evolve to support stakeholder requirements, in particular those represented in the NERC, NCAS and NCEO strategies.
- Develop systems to support Data Citation, and encourage the community to exploit this mechanism to reward data creators, and identify important assets.

#### **Management Practice**

- Maintain a risk register
- Tabulate, and review all activities via appropriate information systems.
- Communicate CEDA activities via an annual report to the community (this report)
- Monitor performance, via the collection, analysis and dissemination of existing and new metrics.
- Survey “data user” and “data producer” communities regularly and analyse results.



# CEDA 2008/2009 Report

## Introduction

In this section we present the report for the 2008/2009 financial year (April to April). The report consists of four sections: a list of “in-year goals”, some notable events, a series of highlights, and conclude with a set of metrics.

## 2008-2009 Goals

As this is the first annual report in this format, the goals listed here are those which are either in addition to those listed in the general objectives, or which provide additional details, and the arrangement is mostly historical. The list is also somewhat incomplete, omitting most activities which are not funded by, or related to, the NERC service level agreement. In the future we will try to make the “granularity” more appropriate for an annual report, with high level major goals appearing here (supplementing finer level goals appearing in the internal project management system), covering major contracts and actions worthy of report.

## Operations & Management

1. Maintain a fully searchable catalogue of all data holdings.
2. Maintain, run and improve the NERC data discovery service.
3. Ongoing maintenance of, and user support for, existing archived datasets.
4. Ensure reliable data preservation and curation by maintaining (or registering elsewhere) duplicate copies of data and metadata, onsite and offsite as deemed necessary and listed in the risk register.
5. Maintain “semi-operational” data ingestion from Met Office, ESA, Eumetsat and ECMWF.
6. Develop more resilient hardware environment so as to minimise system downtime.

## Science Support

1. Ingest new data as required by community and provide community support.
2. Acquire and document data required by the emerging NCEO theme requirements.
3. Acquire and put in place ongoing support for, new CRU analysis dataset.
4. Where appropriate, provide services for the management of intermediate data and information products, enabling NCAS and NCEO scientists to exchange intermediate results.
5. Monitor compliance with the NERC Data Policy.
6. Continue to acquire data from historical NERC funded programmes and liaise with their communities (e.g. Polluted Troposphere).
7. Carry out data management tasks for: QUEST, SOLAS, FREE, APPRAISE, CASCADE, COPS< ABACUS-IPY, COBRA-IPY, RAPID-WATCH, VOCLAS and Amazonica.
8. Contribute to the development of the Metafor Common Information Model.
9. Support data sets produced by aircraft (e.g. EuFAR, FAAM, ASRF).

## Development

7. Deploy capabilities of the Big Data Analysis Network in support of the archive.
8. Update information management systems to capture metadata, and make them available to data maintainers and users alike.
9. Develop catalogues that are interoperable with key national and international data cataloguing activities (e.g. INSPIRE<sup>2</sup>, and GEOSS<sup>3</sup>)
10. Develop prototype web services with the ESA Service Support Environment.
11. Streamline ingestion systems to cope with increased volumes of data, and improve metadata..

---

<sup>2</sup>INSPIRE:

<sup>3</sup>GEOSS:



12. Interim support for the NERC DataGrid: Develop new MOLES, upgrade discovery service and interface, improve vocabulary service, improve data services, and develop future roadmap.
13. Develop and access control system for Metafor and CMIP5.
14. Finish the OMII-UK upgrade to the NDG security software.

## Research

15. Govern and manage appropriate controlled vocabularies and ontologies, in particular, the standard names within the Climate Forecasting conventions for NetCDF.

## Notable events

On 3rd December 2008, BADC signed an MoU with the Programme for Climate Model Diagnosis and Intercomparison (PCMDI), and the World Data Centre for Climate (WDCC), on how the three parties will work together to support the production of the next (fifth) IPCC assessment report (AR5), due to be published in 2013.

A document repository was created by CEDA to store reports, posters, images and other unpublished research outputs. The aim was to provide a subject-based repository for the meteorological sciences. It has been released and advertised to BADC and NEODC users. The service continues to be back populated with information already held in the BADC and NEODC data centres. The repository currently has 251 items and has 19 registered users for deposit.

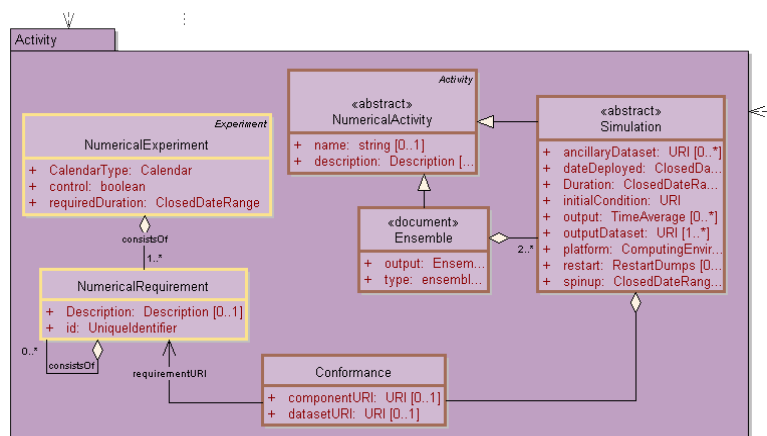
CEDA has recruited two new members of staff in to fill posts which were left vacant at the end of last year. Jurgen Reinecke joined the team at the end of January to manage aircraft data from FAAM and ARSF. Daniel Hagon joined at the start of January as the storage coordinator, whose duties include ensuring data integrity is maintained via the planning and deployment of appropriate storage solutions.

## New Projects Started

### Metafor

*(a project funded by the Framework 7 programme of the European Commission)*

The main objective of METAFOR is to develop and deploy a Common Information Model (CIM) to describe climate data and the models that produce it in a standard way, and to ensure the wide adoption of the CIM. METAFOR will address the fragmentation and gaps in availability of metadata (data describing data) as well as duplication of information collection and problems of identifying, accessing or using climate data that are currently found in existing repositories. METAFOR will optimize the way climate data infrastructures are used to store knowledge, thereby adding value to primary research data and information, and providing an essential asset for the numerous stakeholders actively engaged in climate change issues (policy, research, impacts, mitigation, private sector).



**Figure 1: The METAFOR CIM: A simulation is a numerical activity which conforms to the numerical requirements of a numerical experiment.**

CEDA is heavily involved, providing technical leadership, software engineering, data scientist support, contract management, and more.

The project runs for three years, terminating in March 2011.



## **CORRAL (UK Colonial Registers and Royal Navy Logbooks)**

*(a short project funded by the Joint Information Systems Committee, JISC)*

The CORRAL project imaged 300 logbooks in the special collection of ‘voyages of discovery’ (catalogued under ADM55 in TNA) and from ships on hydrographic surveying service (ADM 53), producing over 50,000 high-quality images all of which are freely available on the BADC website<sup>4</sup>. These images are catalogued by officer, ship, period of voyage and destination(s) thereby allowing users to quickly identify those items that might be of particular interest. Weather observations from published secondary sources of some of these voyages are also available from work done under the international ACRE initiative, and have been used to supplement the primary sources when needed in the visualisations of particular voyages.



## **Infrastructure for the European Network for Earth System Modelling, IS-ENES**

*(a project funded by the Framework 7 programme of the European Commission)*

IS-ENES brings together 6 Earth system modelling groups and two archives (BADC and the World Data Centre for Climate) to establish a European Infrastructure for Earth System Modelling. BADC is involved in two work packages leading one to develop the software for distributed data archives and co-leading a related work package to implement the services. This work is closely coupled with the METAFOR project which is developing the cataloguing information model to describe the data in the archive. The initial phase of IS-ENES is focussed on delivering the European component of the CMIP5 archive for the fifth IPCC Assessment Report. The CMIP5 archive will be more than 100 times larger than its predecessor (the CMIP3 archive used in the fourth IPCC Assessment Report) and will have a huge political significance. The data management work packages of the IS-ENES project will ensure that there are advanced software tools in place to give scientists easy and flexible access to the huge volumes of data in the CMIP5 archive.

The project began in March 2009 and runs for five years.

## **Projects Completed**

### **Overlay Journal Infrastructure for the Meteorological Sciences – OJMS**

*(a project funded by the Joint Information Systems Committee, JISC with complementary funding from NCAS and direct from NERC).*

Overlay journals utilise a technology which is already being used to facilitate peer review and publication on-line. The OJIMS project aimed to investigate this technology further, looking into the business cases and requirements needed to create and operate a Journal of Meteorological Data, where datasets could be published and cited. This would benefit the data scientists creating the datasets by giving them academic recognition for their work, while ensuring that the resulting datasets were of high quality and would be stored and curated in recognised data centres, allowing more reuse of the data.

The OJIMS project was completed in April 2009, and concluded from interactions with meteorological and atmospheric data scientists and organisations that there is a strong need for a data journal. It is hoped that this will be done under the auspices of the Royal Meteorological Society.

<sup>4</sup> The corral data is at <http://badc.nerc.ac.uk/data/corral/index.html>.



## Collaborations

In 2008/2009, significant national and international collaborations have been continued and/or begun. On the national and international scale, NCAS/BADC and NCEO/NEODC a

1. Other NERC centres, under the auspices of interim NERC funding to continue the NERC DataGrid work,
2. The Royal Meteorological Society (RMetS), via JISC and NCAS funding to prototype infrastructure to support the establishment of a new “data journal” to be managed by the RMetS.
3. The US Programme for Climate Model Diagnosis and Intercomparison and their Earth System Grid partners (particularly those at NCAR and GFDL) on software to support the forthcoming fifth Coupled Model Intercomparison Project (CMIP5).
4. PCMDI and the World Data Centre for Climate (WDCC, at the Max Planck Institute for Meteorology, Hamburg) on delivering a globally distributed archive to support CMIP5 (an MoU was signed in December),
5. The International Coupled Chemistry Climate Modelling Validation project (CCMVAL),
6. European partners in the EC funded Metafor consortium to document climate models, particularly in the context of model intercomparison projects such as CMIP5,
7. The IPCC to deliver the IPCC-data centre ([www.ipcc-data.org](http://www.ipcc-data.org)), with Defra funding, and in partnership with WDCC and CIESEN in New York.
8. The Met Office, on many fronts, and with a variety of funding, but particularly in the context of two further Defra funded activities:
  - a. The UK Climate Impacts Link Project, and
  - b. The delivery of the portal for the UKCP09 portal to probabilistic climate projections supplied by the Met Office and delivered in partnership with the UK Climate Impacts Programme.
9. The European Space Agency (ESA) on
  - a. development and operation of a processor and archive for (A)ATSR data products
  - b. development of ESA Service Support Environment (SSE) web services allowing users to extract and manipulate datasets on demand
10. The wider UK atmospheric science and Earth Observation community, via a range of projects, with NCAS, NCEO and other NERC funding

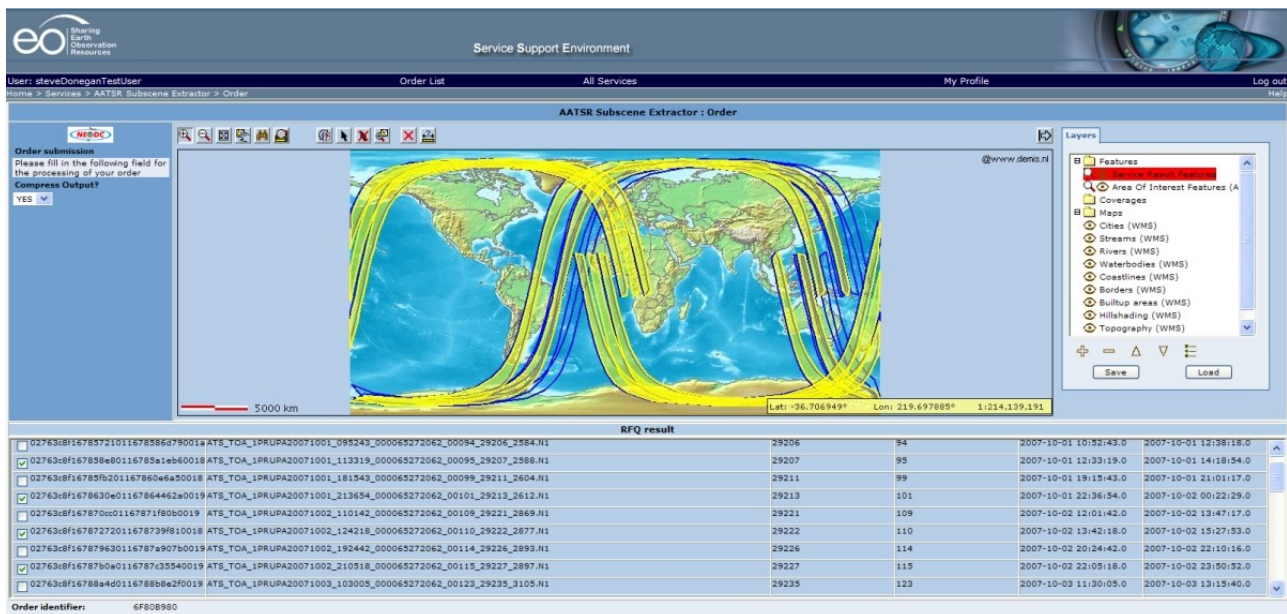


## Highlight: ESA Service Support Environment and Heterogeneous Mission Accessibility

Dr S. Donegan, Dr M. Pritchard, Dr B. Lawrence, Dr V. Jay and Dr A. Woolf<sup>5</sup>

The ESA Service Support Environment (SSE) provides a set of generic components for interoperable web services, and staff at NEODC have worked under ESA funding to create a number of web services using the SSE. Three NEODC services have most recently been developed; two services to produce derivative products from (A)ATSR, with the aim to provide tools for NEODC users to easily extract and manipulate data in the NEODC (A)ATSR archive, and a further service to enable SSE users to discover NERC (and other) data holdings using the NERC Data Discovery Service.

The AATSR services are a Subscene Extractor (see figure below), and Radiance Log which allows users to generate a radiance/reflectance log for any location, e.g. for plotting radiance trends for volcanoes or in calibration/validation activities. The services are available to registered users of the ESA SSE Portal and NEODC.



NEODC SSE Service allows users to specify a geographic subscene and timescale, then to extract this area from all satisfactory AATSR archive data before downloading the data

The SSE now forms part of the prototype infrastructure for the ESA Heterogeneous Missions Accessibility (HMA) architecture, a major strategic development enabling diverse EO ground segments to interoperate with each other and with catalogues and client applications. The HMA together with the SSE, is key to future technical interaction with ESA (and other agencies), so a strong capability in this area is crucial to ensure the future availability of ESA datasets and services to the NEODC/NERC community.

CEDA participation in HMA-activities and development of standards continues with a recent successful bid to ESA to extend existing EO Product Metadata schema for more EO products and instrument types under the HMA Follow On project, as well as involvement in the ESA Long Term Data Preservation Programme.

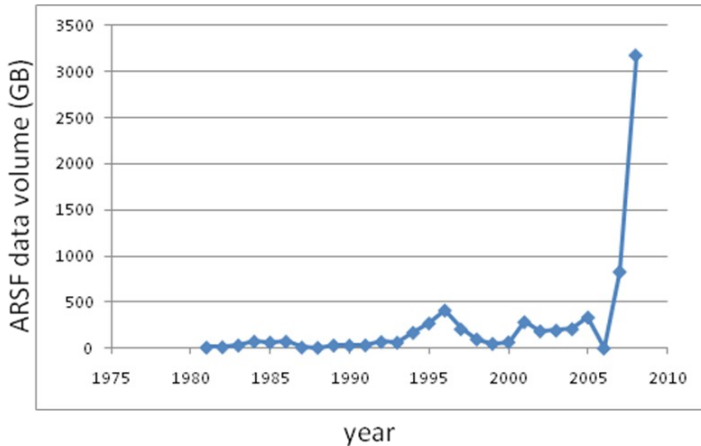
<sup>5</sup> STFC e-Science



## Highlight: Acquire, curate and provide access to Earth observation data generated by NERC funded activities: ARSF (Airborne Research and Survey Facility)

Dr V. Jay, Mr J. Reinecke, Dr S. Donegan, Dr M. Pritchard

NEODC archives and makes available over 25 years' of NERC ARSF data, including aerial photographs, multispectral imagery and LiDAR data. The data are currently processed by ARSF-DAN, the ARSF Data Analysis Node at PML Plymouth, from where they are transferred to NEODC for long-term archival and dissemination. This



year has seen a significant increase in data volumes routinely acquired and archived, with the introduction of the new SPECIM Eagle and Hawk hyperspectral instruments.

ARSF data is the most popular dataset at NEODC, with 41 users downloading digitised aerial photographs (90GB), and 50 users downloading over 200GB of data from the other ARSF instruments: ATM, CASI, LiDAR, Eagle and Hawk in 08/09.

NEODC staff regularly provide guidance and advice to the ARSF operations and processing teams on data related matters. Victoria Jay is a member of the ARSF steering committee and NEODC participates in ARSF user workshops. A formal Data Management Plan for ARSF was produced for the first time this year in consultation with the ARSF Operations and Data Processing Teams.

One activity this year involved the “rescue” of old ARSF data previously held at CEH Monks Wood. NEODC collected approximately 25 boxes containing 1500 CDs, 700 DAT tapes, 600 Exabyte tapes. This included raw and processed data, calibration and



navigation data and miscellaneous back-ups. These items filled a number of gaps in the NEODC archive and all processed data has now been ingested and catalogued.



ARSF multispectral image data (ATM, CASI, Eagle, Hawk) archived since 1995 is in HDF format, with metadata included in the files. This allows software to automatically extract data locations, e.g. for display on the interactive map. Work has begun this year on processing pre-1995 ATM data in the NEODC archive so it can be added to the on-line archive in HDF format including appropriate metadata.



## Highlight: Services in support of NCEO and the wider environmental science community

### Example 1: Eumetsat MetOp data for NCEO Science Themes

**Dr V. Jay, Mr D. Hagon, Mr A. Harwood, Dr S. Pepler**

The National Centre for Earth Observation incorporates 7 Science Themes with over 120 scientists in more than 30 working on a wide range of EO related projects. NEODC, as the designated data centre for NCEO, have over the course of the year established contact with the different Themes and carried out an initial assessment of datasets to be produced (which will require long term archival at NEODC), and third party datasets required (which may be made available through NEODC).

A common requirement to several themes was easy on-line access Eumetsat MetOp data (IASI, GOME-2 and AVHRR-3 instruments). The large data volumes involved required bulk orders of data on 800GB LTO-4 tapes, which have been ingested and made available on-line to registered users at NEODC.



The MetOp data archive at NEODC will initially hold 16 TB IASI (Level 2 and Level 1b) data, 6 TB AVHRR-3 (Level 1b) data and 4 TB GOME-2 (Level 1b) data covering 2007 and 2008, and will continue to grow as data from the ongoing and future EPS-MetOp missions are added.



### Example 2: Data archive for NCAVEO field experiment

**Dr V. Jay, Dr S. Donegan, Dr M. Pritchard**

NCAVEO (Network for Calibration and Validation in Earth Observation) is a NERC funded knowledge transfer network, which organised a community-led field campaign based at Chilbolton in June 2006.



The campaign involved 37 scientists from 22 UK and international organisations (academia, industry and government). The experiment included the collection of data from seven EO satellites, three aircraft and field measurements from several teams on the ground.

NEODC was responsible for archiving and disseminating the data, and will preserve the data for the long term. The complete NCAVEO Field Campaign data archive holds data from over 25 instruments and is over 100 GB in size.

During and after the experiment, NEODC worked with the community on format and metadata issues and more generally educated the scientists involved about data management.

*Milton, E.J. and NCAVEO Partnership, (2008) The Network for Calibration and Validation in Earth Observation (NCAVEO) 2006 field campaign. In, Proceedings of the Remote Sensing and Photogrammetry Society Conference: "Measuring Change in the Earth System". University of Exeter, 15-17 September 2008. Nottingham, UK, Remote Sensing & Photogrammetry Society, 1-4.*

## Highlight: The (A)ATSR archive at NEODC and associated activities

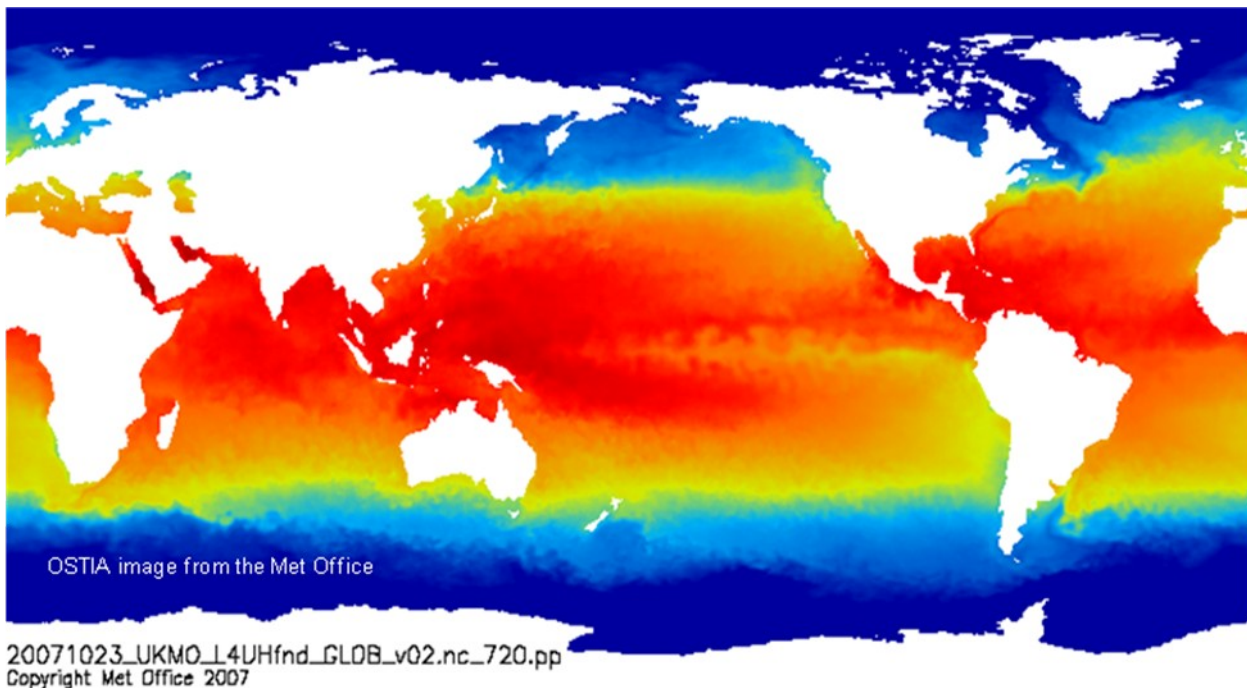
Dr M. Pritchard, Mr D. Grant<sup>6</sup>, Mr B. Maddison<sup>7</sup>

The first ATSR (Along Track Scanning Radiometer) was a research instrument designed and built at the Rutherford Appleton Laboratory (RAL), in the UK and was launched in 1991 on board ESA's ERS-1 satellite. The ATSR instrument series (ATSR-1, ATSR-2, AATSR on board ERS-1, ERS-2 and ENVISAT satellites) have since provided a unique set of instruments tailor-made for high-accuracy measurement of sea-surface temperature (SST) from space.

The partners in the ATSR project have a common aim to encourage the use of (A)ATSR data, to facilitate access to the dataset and to maximise its exploitation. As a result, the (A)ATSR Archive has been successfully providing users with ENVISAT-format data products from AATSR as well as ATSR-1 and ATSR-2 for the past 4 years. The archive is jointly hosted by ESA and the NEODC.

Recent projects, funded by ESA and Defra, have involved staff from NEODC/CEDA and have led to the availability of a number of new datasets:

1. products from the recent ESA reprocessing of AATSR,
2. products from the recent RAL/NEODC reprocessing of ATSR-1 and ATSR-2, and,
3. GHRSSST L2P products generated from all 3 missions.



Improvements to the processor have been used to produce a new dataset covering all 3 missions in GHRSSST L2P format, a community-driven SST format combining SST observations with contextual information from contemporaneous ancillary data sources such as local wind speed and error statistics. Together, these enable analysis systems such as the UK Met Office OSTIA system to make use of the 17-year time series now provided by the (A)ATSR series of instruments.

*Pritchard, M & Houghton, N. New products available from the (A)ATSR Archive. Proceedings of 2nd MERIS - (A)ATSR Workshop, ESA/ESRIN, September 2008.*

<sup>6</sup> Tessella Support Services

<sup>7</sup> STFC SSTD/Space Engineering and Technology Division



## Highlight: Supporting the data archive for CMIP5

**Dr. B. Lawrence, Dr. S. Pepler, Dr. C. Pascoe, Dr. S.A. Callaghan, Dr. S. Pascoe, Mr D. Hagon, Mr A. Stephens, Dr. M. Juckes** (all BADC) and colleagues from: Max Planck Institute for Meteorology, Germany; CICS, Princeton University, USA; PCMDI, USA.

The Intergovernmental panel on climate change (IPCC) has produced four assessment reports (AR) identifying climate change issues in three broad areas: the physical basis of climate change (compiled by working group one, WG1), impacts and adaptation (WG2) and mitigation strategies (WG3). The activities of WG1 were greatly aided in AR4 (2007) by the unique archive of model simulations based at the Program for Climate Model Diagnosis and Intercomparison (PCMDI) at Lawrence Livermore National Laboratory (LNL) in the USA. The AR4 archive was used for hundreds of papers that were analysed during the writing of AR4 itself, and is still being used to produce new peer reviewed publications. A fifth assessment report, AR5, is anticipated for 2013. A key part of this will be the results of the international Coupled Model Intercomparison Project (CMIP5). CMIP5 data will be stored in multiple locations, with complete copies of “core” data to be held at least at the BADC, at the World Data Centre for Climate (Max Planck Institute for Meteorology, Hamburg), and at PCMDI.

The screenshot shows the METAFOR web portal interface. At the top, it says "metafor" with the tagline "finding and understanding simulations of past, present and future climate". Below that is "DRAFT CMIP5 Questionnaire". The main heading is "NCAS Climate HIGEM Project". There are instructions on how to add new simulations or experiments. The "Experiments" section has a table with columns for "name" and "simulations", listing various climate experiments with "Add New" buttons. The "Models and Platforms" section has instructions and buttons to "Add a new Model" or "Copy Model from another Centre". It lists "Models associated with NCAS" (HIGEM, HIGEM2) and "Computing platforms associated with NCAS" (HECTOR).

Early prototype of the METAFOR portal for capturing CMIP5 simulation and model descriptions

BADC direct support for CMIP5 and AR5 has already involved the prototype development of a bespoke CMIP5 web portal to acquire model and simulation descriptions which conform to the METAFOR Common Information Model. (METAFOR is a three year EC funded project which began in March 2008, in which both the BADC and CMS have leadership roles). BADC will deploy the live portal later in the year.

The core experimental archive alone is expected to consist of more than 750 TB of data (potentially up to 1 PB), and extra storage will be needed to hold multiple copies of key datasets for performance reasons. This data will be provided by over 20 modelling centres running more than 50 numerical experiments to produce greater than 6500 years of simulation. BADC has procured a first tranche of storage (400 TB), and is working on upgrading CEDA network bandwidth to ensure the best possible download experience for what is expected to be a large community of users. Other major issues associated with CMIP5 data archive

will be suitable data access services and providing efficient and secure access to this data. Within the current year, BADC has taken a leadership role in developing the authorisation and authentication protocols to be used in a global federation, and in devising a European solution to provide data services. The latter will be delivered under the auspices of the recently initiated EC IS-ENES (InfraStructure for a European Network for Earth Simulation).

**B. Lawrence, V. Balaji, S. Hankin, M. Juckes, M. Lautenschlager, D. Middleton, and D Williams, The European contribution to a Global Solution For Access to Climate Model Simulations. EGU General Assembly 2009, Geophysical Research Abstracts, Vol. 11, EGU2009-4231, 2009**

**B. Lawrence, E. Guilyardi, A. Treshansky, and S. Valcke, Models, Metadata and Metafor, EGU General Assembly 2009, Geophysical Research Abstracts, Vol. 11, EGU2009-4231, 2009.**



## Highlight: Best Practice in Digital Curation

Dr S. Pepler<sup>1</sup>, Dr V. Jay<sup>1</sup>, Dr M. Pritchard<sup>1</sup>, Dr K. Marsh<sup>1</sup>, Dr B. Lawrence,<sup>1</sup> Dr A. Woolf<sup>2</sup>

CEDA exists to both facilitate the delivery of the current science programme, and to ensure the availability of data for reuse (along with preservation of data underpinning the scientific record). The latter is a multi-stage activity, termed curation, beginning with appropriately formatting and adequately documenting the data at ingestion, proceeding via active management of the storage media itself and migration of the accompanying information as necessary, to ensure that it is always fit for purpose for use by a “designated community” of potential users. CEDA is establishing a reputation for best practice in this area:

- Sam Pepler and Bryan Lawrence are both on the editorial board of the Digital Curation Journal

- Sam Pepler was on the organising committee of the 4<sup>th</sup> International Conference on Digital Curation (Edinburgh, November 2008), and Bryan Lawrence and Sam Pepler gave an invited presentation at that conference on “Costing Metadata”.

- Kevin Marsh contributed to the Digital Curation 101 course run by the National Digital Curation Centre (DCC).

- Sam Pepler gave an invited talk on the career paths of data management staff at the DCC/ Research Information Network (RIN) Research Data Management Forum.

- Dr Andrew Woolf and Dr Matt Pritchard are involved in evaluating the use of the ISO Open Archive Information Standard, together with other relevant standards and the ESA SAFE activity to provide a preservation path for Earth Observation data.

- The Environmental Research Funders Forum (ERFF) is attempting to analyse the UK capability for making environmental measurements with a view to greater coordination. Dr Andrew Woolf and Dr Sam Pepler have been requested to advise them on the best way to coordinate the gathering of records into a single catalogue that can be examined by policy makers.

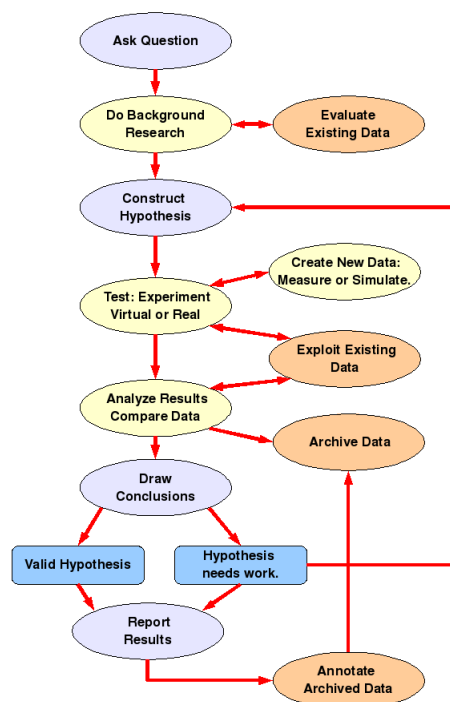
- Dr Andrew Woolf has been involved in testing the ESA Heterogeneous Mission Archive systems user management protocols (within the ESAgrid processing on demand system).

This involvement in the broader curation picture informs and improves practices within CEDA, and the NCAS community. Within the current year this had led to, amongst other things,

- A new cost model for science support activities developed to give quick feedback to upcoming projects and programmes on data management costs.
- A new Data Management Plan template, developed with the involvement of the DCC.
- A new formal internal format review process in order to establish the optimal data encoding for data.
- Kevin Marsh giving a training session at the NCAS summer school on Arran. This was an opportunity to school PhD students in data management skills.

*“Metrics for Success in the Preservation of Scientific Data at the STFC Centre for Environmental Data Archival (CEDA)”. EGU General Assembly 2009, Geophysical Research Abstracts, Vol. 11, EGU2009-4231, 2009*

<sup>1</sup>CEDA; <sup>2</sup>STFC e-Science.

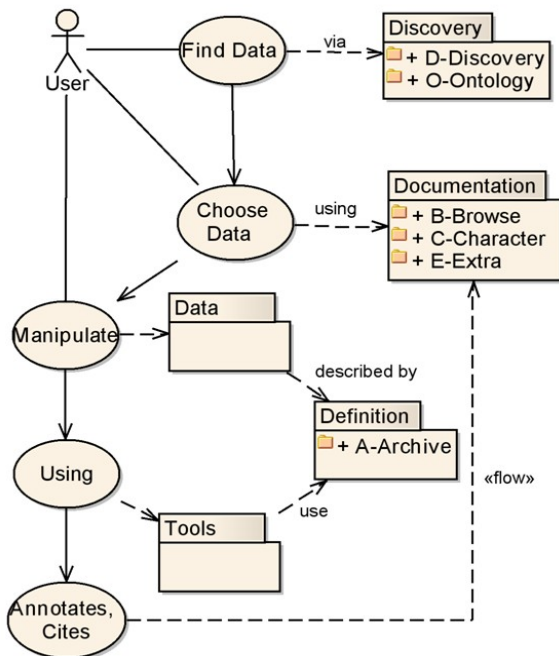


The role of data in the scientific method: A reminder that data curation is about delivery of scientific outcomes, and is not for its own sake.



## Highlight: Information Management Technologies

Dr B. N. Lawrence<sup>1</sup>, Dr S. Cox<sup>2</sup>, Dr S. Donegan<sup>1</sup>, Mr D. Lowe<sup>1</sup>, Dr A. Pamment, Dr M. Pritchard<sup>1</sup>, Dr S. Ventouras<sup>1</sup>, Dr. A. Woolf<sup>3</sup>



Providing homogeneous access (“services”) to the heterogeneous data held within CEDA (and the wider community) requires a robust information paradigm which supports and distinguishes between 1) finding the datasets in the first place (termed “discovery” metadata), 2) choosing between similar datasets (“browse” metadata), 3) post fact annotations of the data such as citations (“character” metadata) and 4) describing the physical format of the data (“archive” metadata).

In the current year CEDA has received funding aimed at sustaining and expanding the “NERC DataGrid” activity in anticipation of a key role in delivering the upcoming NERC information strategy. CEDA continues to deploy and enhance the “NERC Data Discovery Service”, which currently provides discovery metadata for 1278 datasets from across NERC and beyond. Work this year has included major performance enhancements, support for the NERC Portals project, and methods for ranking datasets to improve search responses. Work on the information paradigm to support data browsing has also continued at two levels: within CEDA a completely new (V2.0) implementation of the Metadata Objects for Linking

Environmental Sciences (MOLES) system has been developed and deployed, and will go live to support data scientists and data browsers in mid-2009. At the same time, recognising the wider applicability of MOLES, a wider consortium of stake holders has developed an even newer version (currently 3.2), which exploits the new Observations and Measurements (O&M) ISO draft standard to, in principle, support navigation between datasets found right across the environmental sciences. Development of the Climate Sciences Modelling Language (CSML) is also ongoing (for archive description). This year the major developments have been to include better support for “climatology” attributes, and to prepare for migrating CSML to also be completely O&M compliant.

Cutting across all of these “methodologies” are controlled vocabularies, in their own right, and as part of ontologies. CEDA continues to support Dr Alison Pamment (half time) to manage the standard name component of the international Climate and Forecasting NetCDF conventions. This work has direct benefit to CEDA in helping organise and document data holdings, and is integral to a plethora of projects worldwide. A significant amount of work has also been done on improved software environment for vocabulary management which exploits the NERC vocabulary server based at the British Oceanographic Data Centre.

Andrew Woolf continues to represent CEDA (and NERC) via the British Standards Institute on a range of ISO committees. He is also a key member of OGC and European committees, and via the latter in developing the INSPIRE implementing rules, which will directly impact NERC and CEDA as they become law. This work provides both pull-through of the informatics developments in CEDA into standards and the requirements of those standards on CEDA activities. He is also leading the development of sensor webs, in partnership with the Reading Informatics Research Centre, STFC Energy Research Centre. Work has begun on developing implementations for meteorology observations; with the possible deployments of prototypes to harmonise air quality measurements

**B.N. Lawrence, R. Lowry, P. Miller, H. Snaith, and A. Woolf (2009): Information in environmental data grids. *Phil. Trans. R. Soc. A*, 367, 1003 - 1014. doi:10.1098/rsta.2008.0237.**

<sup>1</sup>CEDA; <sup>2</sup>CSIRO Mining, Australia, <sup>3</sup> STFC e-Science.

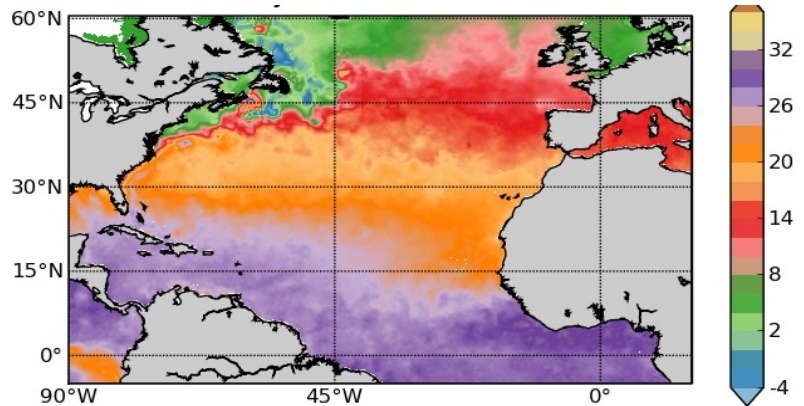
## Highlight: Atmospheric and Earth Observation Research at CEDA

Dr Martin Jukes<sup>1</sup>, Dr Bryan Lawrence<sup>1</sup>, Dr Don Grainger<sup>2</sup>, Jamie Banks<sup>2</sup>

**A New Sea Surface Temperature dataset** is being created using novel new data fusion software created at BADC with previous NOAA and NERC funding.

This work is currently being done as part the AATSR Re-analysis for Climate (ARC) project. BADC will produce new maps of SST with accompanying spatially and temporally resolved uncertainty estimates.

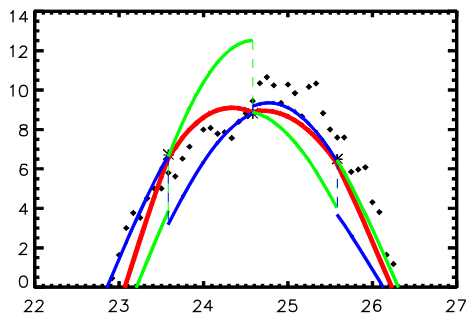
The maps will be at 0.1 degree resolution, capturing fine scale structure of ocean flows.



Sea Surface Temperature [K] for 17th June, 2005, evaluated from reprocessed AATSR data using BADC fusion software.

**Data assimilation theory** has been advanced through a study of the simple harmonic oscillator which shows how information is transferred between

observed and unobserved variables, and how inaccuracies in the assumed model errors feed through into errors in the analysis (Jukes and Lawrence, 2009).

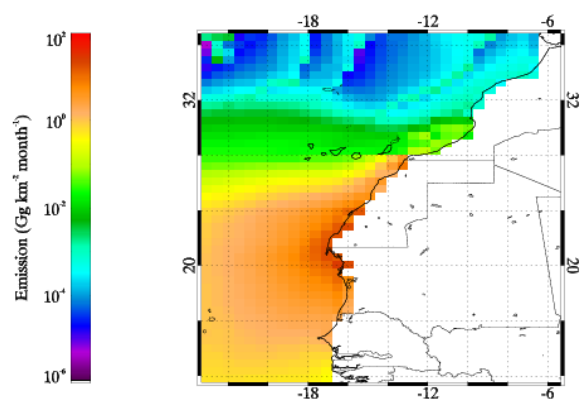


Pseudo observations, Kalman Filter (green), reverse Kalman Filter (blue) and Smoother (red), showing the artificial jumps generated by the Kalman Filter.

**Analysis of Saharan Dust emissions** using data assimilation to constrain the model with observations of aerosol optical depth, provides information about the magnitude and seasonal cycle of aerosol transport from the Sahara and is leading to a better understanding of the dust size distribution.

*Banks, J.R., R. G. Grainger & M. N. Jukes, Lifting and Transport of Saharan Dust, EGU2009-687*

*Jukes M, Lawrence B N (2009) Inferred variables in data assimilation: quantifying sensitivity to inaccurate error statistics. Tellus A 61(1): 129. doi: 10.1111/j.1600-0870.2008.00376.x*



Deposition of Saharan dust in the Atlantic ( $\text{Kg m}^{-2} \text{month}^{-1}$ ) from a model constrained by assimilating satellite retrievals of aerosol optical depth.

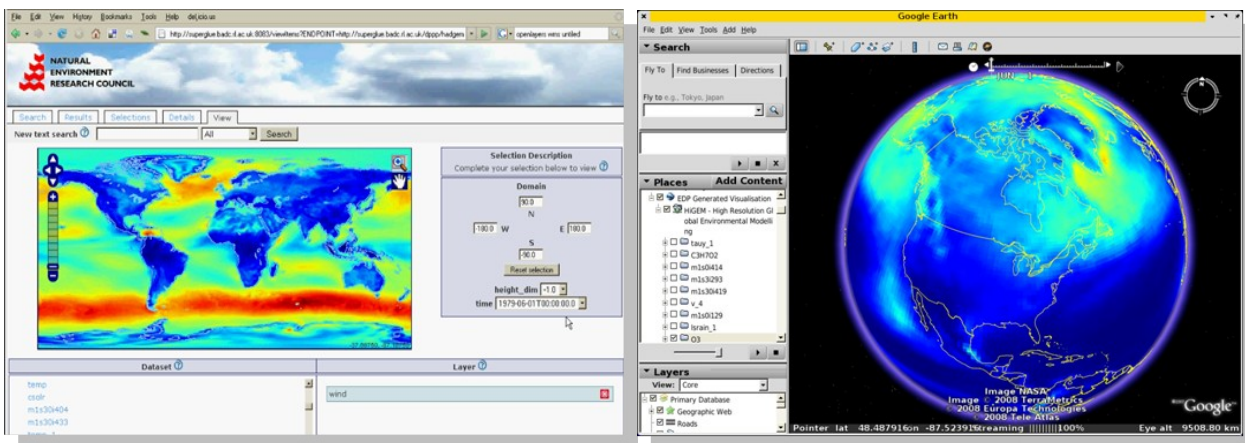
<sup>1</sup>CEDA, <sup>2</sup>University of Oxford



## Highlight: Data Manipulation and Delivery Technologies

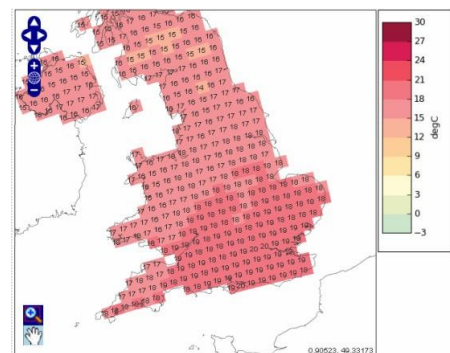
Mr P. Kershaw<sup>1</sup>, Dr B.N. Lawrence<sup>1</sup>, Mr D. Lowe<sup>1</sup>, Dr S. Pascoe<sup>1</sup>, Mr A. Stephens<sup>1</sup>, Dr A. Woolf<sup>2</sup>

A key goal within CEDA is to devise (if necessary) and deploy services which hide the increasing volume and heterogeneity of data from users. Users should be able to extract geographical, temporal and vertical subsets of data according to their requirements in common formats, plot them, and carry out “prescribed” manipulations. There are five key thrusts to the current development work with those aims: developing tools to deliver binary data to users, tools to deliver maps and plots of data to users, tools to deliver structures within datasets (“features”) to users, tools to carry out processing of data for users, and an appropriate access control systems which can work across all these tools. The data access suite will be deployed, along with OPeNDAP services secured with the same access control, to provide access to all suitably formatted CEDA data. Most of the work has revolved around building implementations of the Open Geospatial Consortium web service suite, which provides specifications for Web Coverage Service to deliver data, a Web Map Server for maps, a Web Feature Server for sophisticated subsetting, and a Web Processing Server (WPS) for carrying out data manipulations. Together with portal developments and an access control suite which has been developed in partnership with the U.S. National Centre for Atmospheric Research (in the context of support for CMIP5), a complete data access suite is taking shape.



Two views of the same HIGEM data: one in the prototype portal, and one in Google Earth, both served from the prototype Web Map Server deployed at BADC.

A major effort this year has been developing a user interface (UI) for the Defra UK Climate Projections (UKCP). This “UKCP-UI” provides interactive information on climate change projections to the end of this century. Targeted at policy makers, planning bodies and the general public, it allows users to visualise and retrieve probabilistic climate variables with regional resolution or use a weather generator to simulate weather at their location. The UKCP-UI has been designed for scalability, employing load balancing and job scheduling techniques to withstand a peak of up to 1000 simultaneous users – rather more than the normal requirements expected at CEDA.



CEDA has also developed a prototype which couples a WPS process with a Grid back-end; and in doing so ported a copy of the BADC Trajectory Service to the UK National Grid Service. This work was done as part of an overall scenario around an airport emergency with release of toxic airborne pollutant

*D. Lowe, A. Woolf, B. Lawrence and S. Pascoe (2009): Integrating the Climate Science Modelling Language with geospatial software and services. doi:10.1080/17538940902866161 Int. J. Digital Earth, 2, 29-39.*

<sup>1</sup>CEDA; <sup>2</sup>STFC e-Science.

## Highlight: Overlay Journal Infrastructure for Meteorological Sciences (OJIMS)

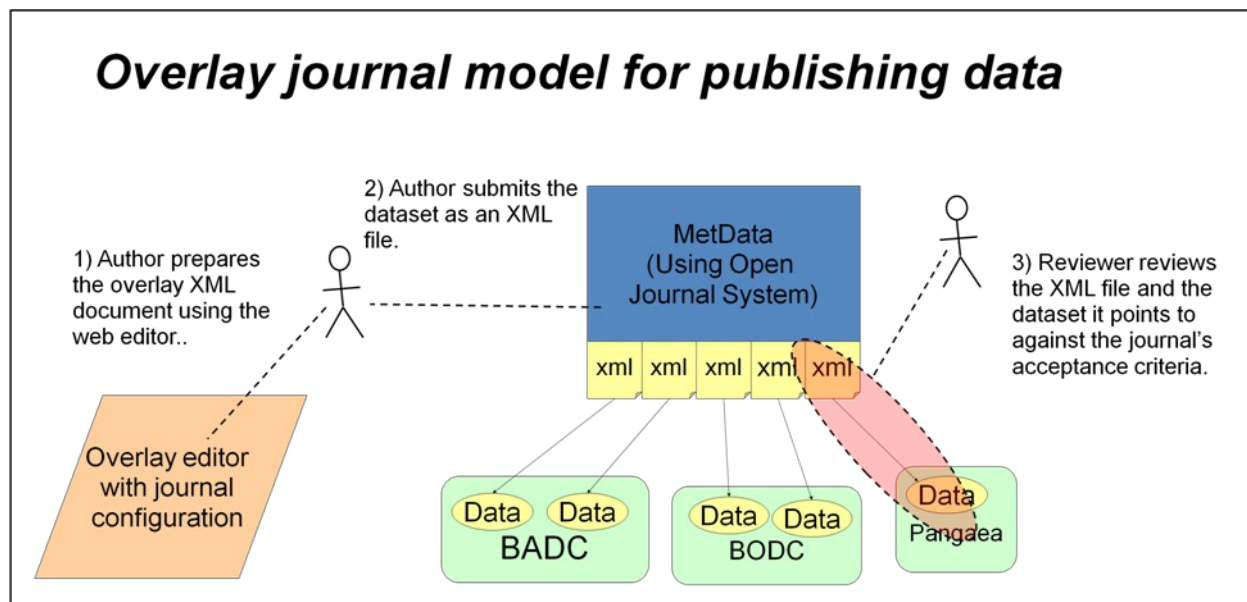
Dr. S. Pepler<sup>1</sup>, Dr. A. Gadian<sup>2</sup>, Dr. S.A.Callaghan<sup>1</sup>, Dr. Paul Hardaker, F. Hewer<sup>3</sup>.

The Overlay Journal Infrastructure for Meteorological Sciences (OJIMS) project has developed mechanisms that could support both a new on-line Journal of Meteorological Data and an Open-Access Repository for documents related to the meteorological sciences.

This JISC and NCAS-funded project aimed to exploit the existing data repository at the BADC along with the expertise of the RMetS to develop the mechanisms to support these activities. This work involved four components:

1. a new open access discipline-specific document repository based at the BADC,
2. the existing BADC data repository,
3. a new overlay journal in which “articles” link peer-reviewed documents to peer-reviewed datasets, and
4. an overlay journal framework that would provide links to highly regarded “kite-marked” (or “star-rated”) papers via the repository (either to the repository contents or the “version of record” held by an original journal publisher).

The project built on the previously JISC funded CLADDIER (<http://claddier.badc.ac.uk>) project and took the next steps towards making these two classes of overlay journal (a “data” journal and a “really useful papers” journal) possible. Interaction with meteorological and atmospheric data scientists and organisations has shown that there is a strong need for a method for publishing data. Publication of data will ensure that the datasets are of good quality, having been peer-reviewed, and will provide data scientists with academic credit for having created the datasets and placing them in an accredited data repository where the data can be archived and curated. Similarly, there is also a desire to have an overlay repository which can serve as a single point of search for numerous institutional repositories.



(JISC is “The Joint Infrastructure Committee” which funds developments and services of benefit to the UK higher educational community.)

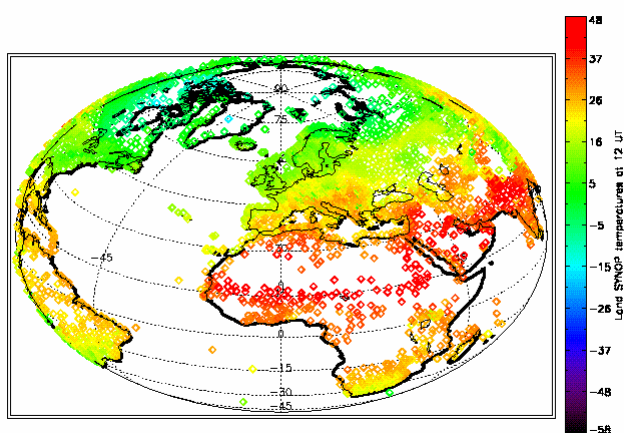
*S.A. Callaghan et al., Overlay Journal Infrastructure for Meteorological Sciences (OJIMS), to be published in Ariadne magazine, 2009.*

<sup>1</sup>BADC; <sup>2</sup> NCAS Weather, Univ. Leeds; <sup>3</sup> Royal Meteorological Society

## Highlight: Examples of Data Acquisition - 1: Met Office MetDB data

Dr G Parton<sup>1</sup>, Dr S Pepler<sup>1</sup>, Mr A Stephens<sup>1</sup>, Dr K Marsh<sup>1</sup>, Dr M Jukes<sup>1</sup>, Stanley Kellet<sup>2</sup>, John Norton<sup>3</sup>, Sheila Needham<sup>2</sup>.

Synoptic meteorological observations find use in a variety of atmospheric research applications: from assisting in deciding where to make new observations, to the evaluation of those observations and of numerical prediction, such data have an important role. NERC researchers have not always had access to such observations, whether required



Land based SYNOP data for XXXX

in real time or not. However, increased liaison between the BADC and the Met Office during the past 12 months has resulted in a number of key observational datasets now being delivered to the BADC. Data including land and ship SYNOP messages, AMDARS reports, global wind profiler data and ozone satellite data. These data arrive up to 4 times a day and often within 6 hours of observation. In addition to the improved timeliness, the acquisition of these data will also ensure the long term preservation of these important data for the atmospheric research community.

Many of the users of such data are not familiar with, or do not have software to handle these data, so in part to address their needs (and those of other users with other

data), a new comma separated values based format has been developed in house (the BADC-CSV format, <http://badc.nerc.ac.uk/help/formats/badc-csv/>). This format will greatly increase the usability of these data to the large section of BADC users who rely on programmes such as Excel to generate and process data.

The new format builds on the functionality of the NASA Ames format, while incorporating important metadata concepts from the NetCDF/Climate Forecasting convention, and the Dublin Core and ISO19115 metadata standards.

## Highlight: Examples of Data Acquisition – 2: the OP3 Campaign.

Mr Andrew Harwood<sup>1</sup>, Dr Spyridon Ventouras<sup>1</sup>, Dr Sam Pepler<sup>1</sup>

The Oxidant and Particle Photochemical Processes above a South-East Asian tropical rain forest (OP3-Danum-08) is a 3-year Consortium Grant of the Natural Environment Research Council (NERC), started 1 October 2007, and as such, the BADC is responsible for its data management. The consortium consists of 23 PIs and co-PIs from eight UK institutions (seven Universities and one NERC laboratory), plus partners from the Malaysian Meteorological Department, University Malaysia Sabah, USA and Europe. OP3 utilized the FAAM aircraft and a Global Atmospheric Watch station (GAW) hosting a 100m research tower, in an undisturbed rainforest in Sabah, Malaysia. Along with providing collaboration tools, BADC staff have provided formatting advice, and assistance with ensuring that the data from 43 different instruments (34 on the ground, and 9 in the air) are appropriately archived and documented at the BADC.



Bukit Atur 100m GAW tower

<sup>1</sup> BADC, <sup>2</sup> Met Office



## Metrics

### Usage – By Quarter

	Apr-Jun, 08	Jul-Sep, 08	Oct-Dec, 08	Jan-Mar, 09
<b>Total Users</b>	11416	11,781	12,289	12,615
<b>Queries Received</b>	921	964	1,084	853
<b>Queries Closed</b>	909	982	1,083	859
<b>Identifiable Users Downloading Data</b>	736; 67	753; 69	754; 77	725; 84
<b>Total Download Volume</b>	7.8; 0.2 TB	11.8; 0.2 TB	3.2; 4.5 TB	10.1; 7.2 TB
<b>Total Number of Files Downloaded</b>	3,100,000; 32,265	4,400,000; 15,993	1,500,000; 513,717	1,800,000; 178,167

(Where appropriate: BADC figures; NEODC figures.)

### Infrastructure

Size of BADC & NEODC primary data copy	256 TB
Total disk capacity	604 TB
Total active NAS devices	35

#### Notable points:

- Over the past year, 1924 BADC users have downloaded 36 TB of data in the form of 12 million files from 134 datasets. 178 NEODC users have downloaded 12 TB of data in 750 thousand files.
- Currently, 69% of registered BADC/NEODC users are from the UK, 12% from other EU countries, and 19% from the rest of the world. 39% of them consider themselves “atmospheric scientists”.
- The BADC/NEODC user base has grown substantially in the past year, rising from 10830 registered users in 2007-2008 to 12615 registered users in 2008-2009.
- The number of user queries has also risen, from 3037 in 2007-2008 to 3418 in 2008-2009 for BADC, and from 322 to 404 for NEODC.
- We have completed our first analysis of the acknowledgements of the use of BADC/NEODC data appearing in the literature. Since 2005, Google Scholar reports that there have been in excess of one hundred acknowledgements per annum of BADC, with 133 in 2008. Acknowledgements of NEODC are growing also, with 20 mentions since 2005. We think these are excellent figures, representing in excess of 500 “citation equivalents” in four years (particularly given that we think many users of BADC/NEODC data do not formally acknowledge the data centre.)



## Funding and Co-Funding (2008/09)

The NCAS British Atmospheric Data Centre and NCEO NERC Earth Observation Data Centre form part of the STFC Centre for Environmental Data Archival (CEDA). The following figures demonstrate that the NCAS contribution is 37% and the NCEO contribution 15% of the overall budget.

NCAS Core Funding: 970K (37%) ; NCEO Core Funding: 378K (15%) ; Co-Funding: 1,249K (48%)

Funding type		income
NCAS BADC	NCAS support for core data centre functions	970
NERC Thematic	Income directly from directed mode programmes	302
NERC NEO DC	NCEO support for core data centre functions	378
NERC other	NERC science and informatics grants	486
Defra	Climate Impacts community development and support	349
JISC	Data publication research	19
EU	METAFOR	50
Other		43
TOTAL		2597

## Publications

Bulgin, C. E., P. I. Palmer, G. E. Thomas, C. P. G. Arnold, E. Campmany, E. Carboni, R. G. Grainger, C. Poulsen, R. Siddans, and B. N. Lawrence (2008), Regional and seasonal variations of the Twomey indirect effect as observed by the ATSR-2 satellite instrument, *Geophys. Res. Lett.*, 35, L02811, doi:10.1029/2007GL031394

D. Lowe, A. Woolf, B.N. Lawrence and S. Pascoe (2009): Integrating the Climate Science Modelling Language with geospatial software and services. doi:10.1080/17538940902866161 *Int.J. Digital Earth*, 2, 29-39.

B.N. Lawrence, R. Lowry, P. Miller, H. Snaith, and A. Woolf (2009): Information in environmental data grids. *Phil. Trans. R. Soc. A*, 367, 1003 - 1014. doi:10.1098/rsta.2008.0237.

S.E. Latham, R. Cramer, M. Grant, P. Kershaw, B.N. Lawrence, R. Lowry, D. Lowe, K O'Neill, P. Miller, S. Pascoe, M. Pritchard, H. Snaith, A. Woolf (2009): The NERC DataGrid Services. *Phil. Trans. R. Soc. A* (2009) 367, 1015- 1019. doi:10.1098/rsta.2008.0238.

M.N. Jukes and B. N. Lawrence (2009): Inferred variables in data assimilation: quantifying sensitivity to inaccurate error statistics. *Tellus A*, 61, 129-143. 10.1111/j.1600-0870.2008.00376.x

Ventouras S., R.Purvinskis, D.Smith and F. Bard, "Propagation effects on non-fixed elevation angle slant paths at frequencies around 26 GHz": in *Radiowave Propagation Models, Tools and Data for Space Systems*, ESA Workshop Publication WP-298, ESA/ESTEC, Noordwijk

Pritchard, M & Houghton, N. New products available from the (A)ATSR Archive. Proceedings of 2nd MERIS - (A)ATSR Workshop, ESA/ESRIN, September 2008.

Milton, E.J. and NCAVEO Partnership, (2008) The Network for Calibration and Validation in Earth Observation (NCAVEO) 2006 field campaign. In, *Proceedings of the Remote Sensing and Photogrammetry Society*



Conference: “Measuring Change in the Earth System”. University of Exeter, 15-17 September 2008. Nottingham, UK, Remote Sensing & Photogrammetry Society, 1-4.

## Presentations

EGU General Assembly Vienna, 2008: Lowe, D.; Woolf, A.; Lawrence, B.N.; Pascoe, S Integrating the Climate Science Modelling Language with geospatial software and services

EGU General Assembly Vienna, 2008: Lawrence, B.N.; Woolf, A.; Lowe, D.; Pascoe, S. Beyond simple features: Do complex feature types need complex service descriptions?

AGU Fall meeting, 2008: Nativi, S., B. Domenico, D. Lowe, E. Davis, P. Mazzetti, Delivering CF-netCDF Coverages via Standard Protocols: a WCS Extension Standard", Eos Trans. AGU, 89(53), Fall Meet. Suppl., Abstract IN33A-1161, 2008.

4th International Digital Curation Conference, Edinburgh, Nov 2008: Lawrence, B.N., S. Pepler, “Costing Metadata”.

Workshop on the use of GIS/OGC standards in meteorology, ECMWF, Nov 2008: Lawrence, B.N., D. Lowe, P. Kershaw and S. Pascoe. Deploying secure OGC services in front of a heterogeneous data archive.

Seminar at the Geophysical Fluid Dynamics Laboratory, Princeton USA, Mar 2009:  
Lawrence, B.N. A torrent of data from CMIP5 is about to arrive. Can the IPCC community cope without new thinking?

RIN (Research Information Network) Data Management Workshop. Manchester, 26-27 November 2008, S.J.Pepler (invited talk on the career paths of data management staff)

ESA Long Term Data Preservation (LTDP) workshop, ESA ESRIN, Italy, 27-28 May 2008, V.L.Jay “BADC/NEODC approach to Long Term Data Preservation and implications of a European LTDP Common Policy”

“Exploiting Geoscience Collections” conference, Geological Society of London, 13<sup>th</sup> May 2008. S.J.Donegan “NEODC Data and Services”

## Public Engagement

Juckles, M., Venue: IEA & Marketforce's 9th Annual Conference WATER 2008 22nd & 23rd October 2008 Radisson SAS Portman, London, Title: Forecasting the future: how will changing weather patterns affect water companies.



## Objectives for 2009/2010

In this section we present the major external drivers expected to impact on CEDA activities, the goals for 2009/2010 for the organisation as a whole, and how the personal objectives of the senior staff address some of those specific goals and the overarching strategic objectives.

### **External Drivers**

This section briefly describes some of the expected major external influences which impact upon CEDA objectives.

#### **Delivery Related (Scientific and Strategic)**

1. CMIP5 data will begin to arrive in the latter part of 2009 and early 2010. Ingestion and documentation will require considerable effort. The volume of data anticipated will stress existing storage, and air conditioning systems.
2. Considerable effort will be required to bring the UKCP09 User Interface online with required functionality for the UKCP launch in (UK) summer 2009.
3. NERC is moving to a new information Strategy, and there may be some consequences, both in terms of internal reprioritisation of effort, and opportunities to exploit new funding (or new activities elsewhere). This work may need to be reflected in actual activities within this year.
4. Both the FREE and QUEST projects are nearing completion and significantly more BADC effort both will be required (particularly given the extra effort required for the QESDI project).
5. Existing Defra contracts are nearing completion and will need to be renegotiated.
6. The European Union EUFAR project will require effort, and we will need to investigate what integration of our activities with EUFAR will entail.
7. Major international programmes and directives are expected to require significant engagement (e.g. INSPIRE, GEO and GMES).

#### **Infrastructural (Financial and Organisational)**

1. STFC will move to the research council shared services scheme in autumn 2009. It is highly likely that initial teething problems will delay normal purchasing, and staff will likely have a significant learning curve getting to grips with new procedures etc.
2. The proposed establishment of a "Space City" at Harwell is bound to impact on CEDA activities, although in 2009/2010 the most likely impact is via the requirement for senior staff to contribute to planning document.

#### **Personnel**

1. Two new staff arrived in 2009, both will need to continue to be supported via lighter workloads and closer supervision during the following year.
2. Two further new staff will be sought, and will need to be inducted and closely supervised on appointment.



# Appendix: CEDA Organisational Structure

The following diagram corresponds to the major functional activities within CEDA:

